

GenCore version 5.1.6  
Copyright (c) 1993-- 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:04:49 ; Search time 26 Seconds

(without alignments)  
3937.814 Million cell updates/sec

Title: AAC393336

Perfect score: 5677  
Sequence: 1 MESSGETAGKPMKNIVPQTC.....IDPFTSNVTGPDILEGGINC 1065

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

PIR-73:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5677	100.0	1065	2 T52054	cellulose synthase
2	3984	70.2	1081	2 T05351	cellulose synthase
3	3811	67.1	1026	2 T51579	cellulose synthase
4	3801	67.0	1065	2 F84649	probable cellulose
5	3628	63.9	1084	2 T08583	cellulose synthase
6	3612.5	63.6	1081	2 T52028	cellulose synthase
7	3590	63.2	1088	2 H84604	probable cellulose
8	3473.5	61.2	974	2 T10797	cellulose synthase
9	3317.5	58.4	958	2 T04870	cellulose synthase
10	2808	49.5	685	2 T10800	cellulose synthase
11	2324	40.9	1181	2 D86157	hypothetical prote
12	2303.5	40.6	1145	2 T51546	cellulose synthase
13	2289.5	40.3	1111	2 T05646	hypothetical prote
14	2221.5	39.1	1036	2 D84741	probable cellulose
15	2185.5	38.5	979	2 C86446	cellulose synthase
16	1931	34.0	583	2 T02209	cellulose synthase
17	1092.5	19.2	757	2 T02561	probable cellulose
18	1069	18.8	712	2 T02552	cellulose synthase
19	1043	18.4	755	2 T02553	cellulose synthase
20	1018.5	17.9	748	2 T02560	cellulose synthase
21	953.5	16.8	727	2 T08918	hypothetical prote
22	948.5	16.7	727	2 T08920	hypothetical prote
23	939.5	16.5	686	2 T08919	hypothetical prote
24	934.5	16.5	686	2 E71417	hypothetical prote
25	849	15.0	828	2 B71417	hypothetical prote
26	441.5	7.8	693	2 AF2275	hypothetical prote
27	305.5	5.4	326	2 T08591	hypothetical prote
28	301	5.3	322	2 T12093	hypothetical prote
29	283.5	5.0	322	2 T08592	hypothetical prote

30	273	4.8	874	2 AB0985	probable polysacch
31	273	4.8	888	2 E91180	probable cellulose
32	272	4.8	692	2 S47754	hypothetical prote
33	272	4.8	888	2 H61511	hypothetical 101.6
34	269	4.7	888	2 G68026	probable cellulose
35	265	4.7	759	2 D70422	cellulose synthase
36	252.5	4.4	1596	2 T31338	cellulose synthase
37	245.5	4.3	729	2 A98320	cellulose synthase
38	245.5	4.3	729	2 AD2963	cellulose synthase
39	245.5	4.3	861	2 I39714	cellulose synthase
40	242.5	4.3	664	2 A95889	cellulose synthase
41	228.5	4.0	754	2 A43735	bcsA protein - Ace
42	196	3.5	723	2 C36963	cellulose synthase
43	178	3.1	654	2 D98350	hypothetical prote
44	178	3.1	654	2 AB2932	beta 1,3 glucan sy
45	173	3.0	768	2 B97083	glycosyltransferase

ALIGNMENTS

RESULT 1									
T52054									
cellulose synthase (EC 2.4.1.-) catalytic subunit [validated] - Arabidopsis thaliana									
C:Species: Arabidopsis thaliana (mouse-ear cress)									
C:Date: 20-Oct-2000 #sequence: revision 20-Oct-2000 #text: change 31-Dec-2000									
C:Accession: T52054									
R:Arioli, T.; Peng, L.; Betzner, A.S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.									
Science 279, 717-720, 1998									
A:Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.									
A:Reference number: 213745; PMID:98111412; PMID:9445479									
A:Accession: T52054									
A>Status: Preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: mRNA									
A:Residues: 1-1065 <ARI>									
A:Cross-references: EMBL:AF027174; PIDN:AAC39336.1									
C:Genetics:									
A>Note: Ath-B									
C:Function:									
A:Description: EC 2.4.1.-; cellulose synthase [validated, PMID:98111412]; involved in									
C:Keywords: glycosyltransferase; hexosyltransferase									
Query Match									
Best local similarity 100.0%; Score 5677; DB 2; Length 1065;									
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	MESSGETAGKPMKNIVPQTC	ICSDNNGKTVGDRFYACDICSFPVCRPCYEYERKDNQ	60					
Db	1	MESSGETAGKPMKNIVPQTC	ICSDNNGKTVGDRFYACDICSFPVCRPCYEYERKDNQ	60					
Qy	61	SCPOCKTRYKRLKSPAI	PGDKDEGLADEGTVERNYPQKEIKSERMLGWLTRGKGEM	120					
Db	61	SCPOCKTRYKRLKSPAI	PGDKDEGLADEGTVERNYPQKEIKSERMLGWLTRGKGEM	120					
Qy	121	GEPOYDKVSHNHL	PRLTSRDTSGEFSAAPEFLSVSTTAGKRLPYSDVNOQSPRR	180					
Db	121	GEPOYDKVSHNHL	PRLTSRDTSGEFSAAPEFLSVSTTAGKRLPYSDVNOQSPRR	180					
Qy	181	IVDPVGLGNVAMKER	VDGWMKKQKNTGPTQASERGVADISTDILDAELANDEA	240					
Db	181	IVDPVGLGNVAMKER	VDGWMKKQKNTGPTQASERGVADISTDILDAELANDEA	240					
Qy	241	RQPLSRKVSIPSSRI	NPYIMRLVILCLFLHYRTNPPVNAFALMLVSVICELPAL	300					
Db	241	RQPLSRKVSIPSSRI	NPYIMRLVILCLFLHYRTNPPVNAFALMLVSVICELPAL	300					
Qy	301	SWIIDPQPKWPPVNR	ETYLDRALRYDREGSPQLAANDIVSVDPKKEPLVTANTVL	360					
Db	301	SWIIDPQPKWPPVNR	ETYLDRALRYDREGSPQLAANDIVSVDPKKEPLVTANTVL	360					
Qy	361	SILAVDYPVDKVS	CVYDDGAAMLSPESLATSEFAKRWPFCKKYSIEPPAPEMYFAK	420					
Db	361	SILAVDYPVDKVS	CVYDDGAAMLSPESLATSEFAKRWPFCKKYSIEPPAPEMYFAK	420					

QY 421 IDYLDKQVTSFVKDRAKREYEERKIRINALVSKALCPEGVWMDGTPMGNNTD 480  
 DB 421 IDYLDKQVTSFVKDRAKREYEERKIRINALVSKALCPEGVWMDGTPMGNNTD 480  
 QY 481 HFGMTQVFLGONGGDAEENELPRLYVYSREKRPFGOHKKAGAMALVRSVAVLTNGPF 540  
 DB 481 HFGMTQVFLGONGGDAEENELPRLYVYSREKRPFGOHKKAGAMALVRSVAVLTNGPF 540  
 QY 541 IINLDCDHYINNKSALREAMCFLMPNLCQVYVQFQRFQDIDKNDYARNNTVFPDI 600  
 DB 541 IINLDCDHYINNKSALREAMCFLMPNLCQVYVQFQRFQDIDKNDYARNNTVFPDI 600  
 QY 601 NLRGLDGIQPVYVGTGCVFNRTALYGPPIKVKHKRPSLSKLGGSKRKNKAKKSS 660  
 DB 601 NLRGLDGIQPVYVGTGCVFNRTALYGPPIKVKHKRPSLSKLGGSKRKNKAKKSS 660  
 QY 661 DKKSGRHTDSTVPVFNLDIEEGVAGFDEDEKALMSQMSLEKRFQGSVAVSTLME 720  
 DB 661 DKKSGRHTDSTVPVFNLDIEEGVAGFDEDEKALMSQMSLEKRFQGSVAVSTLME 720  
 QY 721 NGCVPSATPENLKEAIVHISGVEDKSDMGMEIGMIGSVTEEDILTFKMHARGMRSI 780  
 DB 721 NGCVPSATPENLKEAIVHISGVEDKSDMGMEIGMIGSVTEEDILTFKMHARGMRSI 780  
 QY 781 YCMPKLPAKGSAPINLSDRLNOVLRMALGSVEILFSRHCPIMYGYNGRLKFLERFAYVN 840  
 DB 781 YCMPKLPAKGSAPINLSDRLNOVLRMALGSVEILFSRHCPIMYGYNGRLKFLERFAYVN 840  
 QY 841 TTIYPTSTPLMCTCLAVCLFTNOFIIPQISNATISWFLSFLSTFNTGILMMSGV 900  
 DB 841 TTIYPTSTPLMCTCLAVCLFTNOFIIPQISNATISWFLSFLSTFNTGILMMSGV 900  
 QY 901 GIDEMWRNBOFWVIGVSAHLEFAVQGLKVLGIDTNTVTSKASDEGDEFAELTFKX 960  
 DB 901 GIDEMWRNBOFWVIGVSAHLEFAVQGLKVLGIDTNTVTSKASDEGDEFAELTFKX 960  
 QY 961 TTLILPPTLLIVNLGVAVGVSAVINSYGSGMPLFGKLFPAFWIVHLVYPLFLKGLMR 1020  
 DB 961 TTLILPPTLLIVNLGVAVGVSAVINSYGSGMPLFGKLFPAFWIVHLVYPLFLKGLMR 1020  
 QY 1021 QNRTPYVWMSVLLASIFSLMVRIDPRTSVTGTDLLEGGINC 1065  
 DB 1021 QNRTPYVWMSVLLASIFSLMVRIDPRTSVTGTDLLEGGINC 1065

RESULT 2

T05351

cellulose synthase (EC 2.4.1.-) catalytic chain RSM1 - Arabidopsis thaliana

N:Alternate names: protein F8B4.110

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999

C:Accession: T05351

R:Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; De

submitted to the Protein Sequence Database, February 1999

A:Reference number: Z15409

A:Accession: T05351

A:Molecule type: DNA

A:Residues: 1-1081 <BEV>

A:Cross-references: EMBL:AL034567

A:Experimental source: cultivar Columbia; BAC clone F8B4

C:Genetics:

A:Gene: RSM1

A:Map position: 4

A:Introns: 27/3; 93/1; 150/3; 189/1; 252/2; 341/2; 456/3; 502/3; 544/3; 615/3; 704/1; 76

A:Note: F8B4.110

C:Keywords: glycosyltransferase; hexosyltransferase; P-loop

Query Match 70.2%; Score 3984; DB 2; Length 1081;  
 Best Local Similarity 69.7%; Pred. No. 4.6e-307;  
 Matches 741; Conservative 128; Mismatches 158; Indels 36; Gaps 13;

QY 2 ESEGTAGKPMKNIVPQCICSDNVKTVGDCRFVACDICSFPVCRPCRYEERKDGNS 61  
 DB 23 ESDGT - KPLNMNQICQICGDDVGLAETGADVACNECAFPCRPCRYEERKDGTC 80  
 QY 62 CPQCKTRKRLKGAIPEDKDEDLADGTYEFPVPOKEKISEMGLMHLRGGEENG 121  
 DB 81 CPQCKTRFRHRGSPVEGEDEDDV-DIENEFPYAO-----GANKARHORHG 128  
 QY 122 EPQYDKREVSNH--LPRLTSRODTSGEFSASPERLSYST--IAGCKRLPYSS--DV 173  
 DB 129 E-EFSSSSHSESQPIPLHLGHFTVGEI--RIPDQSVATTSGLPGSPDRMNSSPYIDP 185  
 QY 174 NQSPNRRIYD-----VGLGVANKERYDKMKOEKNTGSPVSOASERGVVIDAST 227  
 DB 186 RQPVPRVRIYDPSKDNLSYGLGVNDMKEREKGLKQEKMLQKMLQKYEKGKG-EIE-GT 243  
 QY 228 DILADEALLNDEARQPLSKRVSIPSSRIINPVMVILRLVILCFLLHRTIPVNAFPL 287  
 DB 244 GSNGBELQADDTRLPMKSRVPISSRLPVRVILRLILCFLLQYRTHPVNAFPL 303  
 QY 288 WLVSVCIEIWFALSNILDOFPKMPVFNRETYLDRLALRYDRGEESQLAADIFVSTYDP 347  
 DB 304 WLVSVCIEIWFALSNILDOFPKMPVFNRETYLDRLALRYDRGEESQLAADIFVSTYDP 363  
 QY 348 LKEPPLVANTVLSIADYPVDKVSCYVDDGAAMLSFESLAETSEPARKVPCKKYS 407  
 DB 364 LKEPPLVANTVLSIADYPVDKVSCYVDDGAAMLSFESLAETSEPARKVPCKKYS 423  
 QY 408 IEPRAPEWFAKTIYDKQVTSFVKDRAKREYEERKIRINALVSKALCPEGVW 467  
 DB 424 IEPRAPEWFAKTIYDKQVTSFVKDRAKREYEERKIRINALVSKALCPEGVW 483  
 QY 468 QDGTWMPGNNTGDHPGMIOVFLGONGGLDAEENELPRLYVYSREKRPFGOHKKAGAMA 527  
 DB 484 QDGTWMPGNNTGDHPGMIOVFLGONGGLDAEENELPRLYVYSREKRPFGOHKKAGAMA 543  
 QY 528 LVRSVAVLTNGPFLNLDGDHYINNKSALREAMCFLMPNLCQVYVQFQRFQDIDKN 587  
 DB 544 LVRSVAVLTNGPFLNLDGDHYINNKSALREAMCFLMPNLCQVYVQFQRFQDIDKN 603  
 QY 588 DRYANRNTPVFDINLRGLDGIQPVYVGTGCVFNRTALYGPPIKVKHKRPSLSKLGG 647  
 DB 604 DRYANRNTPVFDINLRGLDGIQPVYVGTGCVFNRTALYGPPIKVKHKRPSLSKLGG 663  
 QY 648 GSRKKNKAKKESDKKKS-RHTDSTVPVFNLDIEEGVAGFDEDEKALMSQMSLEK 706  
 DB 664 GSRKKNKAKKESDKKKS-RHTDSTVPVFNLDIEEGVAGFDEDEKALMSQMSLEK 721  
 QY 707 FQGSVAVTAMENGVPSATPENLKEAIVHISGVEDKSDMGMEIGMIGSVTEEDI 766  
 DB 722 FQGSVAVTAMENGVPSATPENLKEAIVHISGVEDKSDMGMEIGMIGSVTEEDI 781  
 QY 767 LFGPMKHARGMSIYICMPKLPKGSAPINLSDRLNOVLRMALGSVEILFSRHCPIMYGY 826  
 DB 782 LFGPMKHARGMSIYICMPKLPKGSAPINLSDRLNOVLRMALGSVEILFSRHCPIMYGY 841  
 QY 827 NGRLLKFLERFAVNTIYPTISIPLMCTCLAVCLFTNOFIIPQISNATISWFLSFLS 886  
 DB 842 NGRLLKFLERFAVNTIYPTISIPLMCTCLAVCLFTNOFIIPQISNATISWFLSFLS 901  
 QY 887 IFATGILMRSVGVGIDEMWRNBOFWVIGVSAHLEFAVQGLKVLGIDTNTVTSKAS 946  
 DB 902 IFATGILMRSVGVGIDEMWRNBOFWVIGVSAHLEFAVQGLKVLGIDTNTVTSKAS 961  
 QY 947 DEDGPAELIYLFKMTLLIPPTLLIVNLGVAVGVSAVINSYGSGMPLFGKLFPAFW 1006  
 DB 962 DEDGPAELIYLFKMTLLIPPTLLIVNLGVAVGVSAVINSYGSGMPLFGKLFPAFW 1021  
 QY 1007 IYHLVPLKGLMGRNRTPTIYVWMSVLLASIFSLMVRIDP 1049  
 DB 1022 IYHLVPLKGLMGRNRTPTIYVWMSVLLASIFSLMVRIDP 1064



QY	339	IEFVSTVDPIKEPPLVNTANTVSTILIAVDPVQVSCVYEDDDGAAMLSEFSLATSFEAK	398
Db	342	DVFSTVDPKPEPPLVNTANTVSTILIAVDIPVKVACVYDDDSGAMLTFEELSETAESEK	401
QY	399	WVPECKKSIIEBRAEWEYFAAKIDYLDKDVQVTSFYVDRBRAMKEVEEFAIRIALVSKAL	458
Db	402	WVPECKKFNIEBRAEEFYFSQKIDYLDKDIQSFVEXERRAMKEVEEFAVRIMILVAKQ	461
QY	459	KCPREGVWMOOSTPYPRGNNTGDHPGMIQVFLQNGSLDABGNLPLVYVSRKRGFOH	518
Db	462	KIPEDGWTMEDSTSPGNPRDHPGMIQVFLHSGSLDIDGMLPRLIYVSRKRGFOH	521
QY	519	HKKAGAMALVAVSVLTNGPFLINLDDCHYITNNSKALREAMCELMDDPNLQKVCYQFP	578
Db	522	HKKAGAMALIVSAVLNTGAILLVANDCHYFNNSKALREAMCFMMDPAIGKCCYQFP	581
QY	579	QRFEDGIDKNDRYANRNTYFEDINLKGDLGICQPVYVGTGCVENRTALVGEPIKYVHKK	638
Db	582	QRFEDIDLDHRYANRNTYFEDINLGLDIOGPVYVGTGCCFNRLQALVGDPLTFEEDL	641
QY	639	PSLSKTLGGSSKRRKNSKAK-RESDDKKSGRHNDSTVPVNLDDIEGVGAGDDEKALL	697
Db	642	PNLYKSCFGSKKKKSKRIYEDNRSTIKRSDSNPLTFNMEDIDEDVE--GEDEMSLL	699
QY	698	MSQMLEKRFQGSAAVFAVSTLMENGVPVSATPENILKEALIHVISCYEEDKSDWMEIGW	757
Db	700	VSQKRLERFGSPFEIATFMEOGGLPSTPDLTLKKAIHVISCYEAKTDMGKEIGW	759
QY	758	IYGSTEDILTFEKKHARGMRSITCYCPKLPARFGSAPINLSDLNOLVRKALGSVELPS	817
Db	760	IYGSTEDILTFEKKHARGMISITCYPSRPAKRGSAPINLSDLNOLVRKALGSIELLS	819
QY	818	RHCPIMVGYNGRLKELEFAYVNTTYPTISIPLLMYCFLLAVCLEPTNOFIIPQISNIAS	877
Db	820	RHCPIMVGYNGRLKELERAYINTIYVPTTISIPLLAYCMLPACLLTFNFIPEISNIAS	879
QY	878	IMFSLPFIPTGILTEMWSGVGIDEMWRNDFWVIGVSALLFAVPOGLKVLVAGIDT	937
Db	880	LCFMLLEFASIASALLEKWSVALLDEMWRNDFWVIGTSALLFAVPOGLKVLVAGIDT	939
QY	938	NETVYVSKASDEDEGAELVLFKWTYLLIIPPTLLIYNLGVAVAGYAINSGQSMGPLF	997
Db	940	NETVYVSKASDEDEGAELVYFKWTSLLIIPPTLLIYNLGVAVAGYAINSGQSMGPLM	999
QY	998	GKLPFAFYVIALYFPLKGMGRONRPTIYVYVSVLASISBSLWVRIDPFISRYTG	1055
Db	1000	GKLPFAFYVIALYFPLKGLGRONRPTIYVYVSWLASISBSLWVRINPFVS--FTG	1056

## RESULT 5

cellulose synthase (EC 2.4.1.-) catalytic chain - Arabidopsis thaliana  
N.Alternate names: T22F8, T250  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
C.Accession: T08583; T09014  
R.Bevan, M.; Zimmerman, W.; Gruenisen, A.; Wanduit, R.; Bancroft, I.; Mewes, H.W.; May  
submitted to the Protein Sequence Database, May 1999.  
A.Reference number: 216442  
A.Accession: T08583  
A.Molecule type: DNA  
A.Residues: 1-1084 <BEV>  
A.Cross-references: EMBL:AL050351; GSPDB:GN00062; ATSP:T22F8, 250  
A.Experimental source: cultivar Columbia; BAC clone T22F8  
R.Arioli, T.; Peng, L.; Betner, A.S.; Burr, J.; Witke, W.; Herth, W.; Camilleri, C.; H  
science 279, 717-720, 1998  
A.Title: Molecular analysis of cellulose biosynthesis in Arabidopsis.  
A.Reference number: 213745; MUID:98111412; PMID:9445479  
A.Accession: T09014  
A.Status: translated from GB/EMBL/DDBJ  
A.Molecule type: mRNA  
A.Residues: 1-1084 <ARI>

A:cross-references: EMBL:AF027173; NID:q2827140; PID:q2827141  
A:Experimental source: Cultivar Columbia  
C:Genetics:  
A:Gene: ATSP:722F8.250; Ath-A  
A:Map position: 4  
A:introns: 2/3; 156/3; 191/1; 254/2; 343/2; 458/3; 504/3; 546/3; 617/3; 708/1; 773/3  
C:Keywords: cell wall synthesis; glycosyltransferase; hexosyltransferase

Query Match	63.9%;	Score 3628;	DB 2;	Length 1084;
Best Local Similarity	63.0%;	Pred. No. 8e-279;		
Matches 678;	Conservative 165;	Mismatches 191;	Indels 42;	Gaps 16

OY	10	KPKNNIYPOQCIOCSNVKTYDGRFPAACIDISCSPYPCPYEKYERKDQNSOCQCTRY	69
Dd	29	RSVOELSGOTCQICGDEILYTSSELPFAACNECAPVPCPYEKYERKEDGACQCTRY	88
OY	70	KRLGSPAIJPGDAKDEGLADEGTVEFN - POKEKISERMLGMHLTRKGEEMGEPOYKE	128
Dd	89	KRIKSGPRVGDDEDEEDIDLEYEFDHGMDEPHAAEALSSRLNTGRC-----GLDSA	142
OY	129	VSHNHLPRLTSGRODTSGEFSAASPERLVSSTIAGAKL---PSSDYNOSSPNRIYDP-	184
Dd	143	PGSOPJPLLTLYCEDEDWTS--DRHALYPPSTGYNVYPAPE-TDSSAPPOARSWPO	199
OY	185	-----VGLGVAKERVDGMMKMKOKENGPVSTOAAE-RCGVNIDASTOLDLEALND	238
Dd	200	KDIAETGYGSVAKADMEYWKRRQGEKLOYIHEBGGNNGKSSNDDELD---PDMAMD	256
OY	239	EANOPISRKXSPSSRINPYRWIMLRVLICLFPHYRTIPVNAEALMVSYCEIWF	298
Dd	257	EGNOPLSRKPIRSSHINRYMLILCRALILGLFPHYILHPVADAYGLMITSYGEIWF	316
OY	299	ALSMLIDQEPKMEPVNRETYLDRLALRDREGSPQLAANDIEVSTYDPLKEPPLYANT	358
Dd	317	AVSMIDQEPKWPPIRETYLDRLSLREYKEKSPGLGADVDFVSTVDPPLKEPPLTANT	376
OY	359	VLSIADVDPVOKSCYVYFDDCAAMLSPESLAEFSEFARKVNPCKXYSTIEPRAPDEXA	418
Dd	377	VLSIADVDPVDKVCYVSDDGAAMLTEALSDFAEFRKVPCKFNLIEPRAPDEYFS	436
OY	419	AKIDYLKDKQVTSFYVDRBRAMREVEEFKIRINALVSALKCPREGWMDQGTWMPENNT	478
Dd	437	QKMDYLKKNYHPFVBERAKMRDYEEFRVKINALVALAOKVPRBEGTMDQGTWMPENNY	496
OY	479	GDBHGIQVFLGONGGLDAEGNELPRLYYVSREKRPGFQHHKKAGAMNALVRSAYLJNG	538
Dd	497	RDBHGMQVFLGSHGVBDTGDNELPRLYYVSREKRPGEFDHKKAGAMNSILVRSAYLSNA	556
OY	539	PFLINDCCHYINNSKALPEAMCFLMDPRLKQVCYVOPORFEGIDKNDRYANRNTVEF	598
Dd	557	PFLYNDCCDHYINNSKAIRESMCFPMMDPOSGKVCYVOPORFEGIDIRHRYSNRNVFF	616
OY	599	DINLRGDIQGVVYVYGCVFNPRLALXGEBPRIKVKRP-----SLSK---LGGSR	650
Dd	617	DINMGKIDGIQGVYVYGCVFRRQALGFEDEP---KKKRPQGTCCMWKRMCCLOGGLR	673
OY	651	KKN-SKAKESDKKSKSGRHTDSTVEVFNLDIEEGVEGAGHDDDEKALLMSOMSLKRRFO	709
Dd	674	KKSKTKAK--DKTNTKETSQIHA--LEANDGVIVPVSINVEKRSBEOQLKLEKKFGQ	728
OY	710	SAVFVASTMEWNGVPPSATPENLLKEALIHVISCYEDKDSMGMEICMWTIGSYTEDILTG	769
Dd	729	SPVFVASVYLONGVGRNAPSACLLREAIQVYISCGYEDKTEMGKEIMWIGSYTEDILTG	788
OY	770	EKNHAGSRTYCMPLPAFKGSAPILNIDRLNOLYLRALGSGVILFSHRCPITWYGNGR	829
Dd	789	EKNHCGMSVYCMPPRAAFKGSAPILNIDRLHOYLKRALGSGVILFSHRCPITWYGGGG	848
OY	830	LKFLFEPAVANTYITITSIPLMCTCLAAFLNOPRIQOISINISIFLSLSTISIFA	889
Dd	849	LKMLDERSTYNSVYWTSLPIYVCSLPAYCCLLTGRTIYEISINAGILEPMLMEFISAV	908
OY	890	TGILEMRWGSVGIDEMWRNDEQFWJIGVSAHLFAVFOGILKVLAGIDTNEFTVYSKASDE	949



```

Db      264  ROLSKRLPLRSSRINPYRLILFCRLAIIIGLFHYRLILHPVNDAFGLMTLSVCEIMFAV 323
Qy      301  SWLLDPEPKMEPNRETEYLDRLALRYDRGEPSQLAAVDIPEVTPLKPEPLVANTYL 360
Db      324  SWLLDPEPKMYPERETYLDRLSLRKEBGPSELPAVDVETVPLKEPPLITANTYL 363
Qy      361  SILAVDPYDKVSCYVFDGAAMLSESLAETSEFARKWPFCKKYSIEPRAEWFAAK 420
Db      384  SILAVDPYEVKACVYSDGAAMLFEALSTAEFARKWPFCKKYSIEPRAEWFAAK 443
Qy      421  IDYLAKKRVOTSPKDRAMKREYEERKIRINALYSALKCPDEGWMODGTPWPGNNTGD 480
Db      444  MDLAKRVDPAFMEARKRDYEERKIRINALYSALKCPDEGWMODGTPWPGNNTGD 503
Qy      481  HPGMIVPFLGONGDLAEENELPRLYVSREKRPQOHKKAGAMALVAVSVLNGPE 540
Db      504  HPGMIVPFLGSHSVCCMDGNELPRLYVSREKRPQOHKKAGAMALVAVSVLNGPE 563
Qy      541  IINLDCHYINNSKALREAMCFPLADPNLKGQVYVOPQRFDDIDKNDRYANRYTFDDI 600
Db      564  IINVDCHYINNSKALREAMCFMDPQSGKKICYVOPQRFDDIDKNDRYANRYTFDDI 623
Qy      601  NLRGLDIOGPVYVGTGVPNRTALYGEPPKVKHKP-----SLISK---LCGSRKK 652
Db      624  NMKGIDIOGPVYVGTGVPNRTALYGEPPKVKHKP-----SLISK---LCGSRKK 680
Qy      653  NSAKKESDKKSGRTDSTVPVFNLDIIEGVGAGFDEKALIMSOMSLERKQSANV 712
Db      681  KTGKVDONOKRP---KETSKQIHAEHLEEGQVLTNAENNSP---TQALKLEKKEFGQSPV 735
Qy      713  FVASTLEMNGVPPSATPENLLEALHIVISCGEYEDKSDMGMEIGWYIGSTEDILGFKM 772
Db      736  IYASTLLINGGVSNVNPASLRESIQVISCYEETKEMKEIGWYIGSTEDILGFKM 795
Qy      773  HARGMSIYCMPLRPFKGSAPINLSDRLOVLRMALGSVEIILPSRCPITWIGYNGRLK 832
Db      796  HCGMGSYVCMPLRPFKGSAPINLSDRLOVLRMALGSVEIILPSRCPITWIGYNGRLK 855
Qy      833  LEFEAVNTTYITPITSPLMCTLAVCLFTNOFIIPOTSINASIIMPLSLFSTFATGI 892
Db      856  LEFEAVNTTYITPITSPLMCTLAVCLFTNOFIIPOTSINASIIMPLSLFSTFATGI 915
Qy      893  LEMRMSVGIIDEMRNEQFVWJGVSAHLFAVFOGILKVLAGIDTNETYTSKASDEGDF 952
Db      916  LEMRMSVGIIDEMRNEQFVWJGVSAHLFAVFOGILKVLAGIDTNETYTSKASDEGDF 974
Qy      953  AELYLKRWTLILPPTLLIIVNLGVYAVSYAINSGYOSWGLPEKLPFAFVYVHLXP 1012
Db      975  SELYLKRWTLILPPTLLIIVNLGVYAVSYAINSGYOSWGLPEKLPFAFVYVHLXP 1034
Qy      1013  FLKGLMGQRNRPETIVVWVSVLLASIFSLMVRIDPFTSRVTGPDILGEGINC 1065
Db      1035  FLKGLMGQRNRPETIVVWVSVLLASIFSLMVRIDPFTSRVTGPDILGEGINC 1086

RESULT 8
T10797
cellulose synthase (EC 2.4.1.-) catalytic chain celA1 - upland cotton
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10797
R:Pear, J.R.; Kawagoe, Y.; Schreckengost, W.E.; Delmer, D.P.; Stalker, D.M.
Proc. Natl. Acad. Sci. U.S.A. 93, 12637-12642, 1996
A:title: Higher plants contain homologs of the bacterial celA genes encoding the catalytic
A:reference number: Z17152; MIMD:9705726; PMID:8901635
A:Accession: T10797
A:status: preliminary; translated from GB/EMBL/DBJ
A:molecule type: mRNA
A:Residues: 1-974 <DEA>
A:Cross-references: EMBL:058283; NID:g1706955; PION:AB37766.1; PID:g1706956
C:experimental source: strain Acala SJ-2; fiber
C:genetics:

```

```

A:Gene: celA1
C:Function:
A:Description: involved in the synthesis of cellulose
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match      61.2%; Score 3473.5; DB 2; Length 974;
Best local Similarity 62.1%; Pred. No. 1.2e-266;
Matches 660; Conservative 129; Mismatches 176; Indels 97; Gaps 15;

Qy      12  MKNIYVQTOIGSDNKGTVDDGRFVACDICSPPVCPPEYERKGNOSQOQKRYR 71
Db      2  MESGP-VCHTGEHGLVNGEPFVACHENPICKSCFEDLKGKACLRG----- 54
Qy      72  LKSGPAPIDKDEDGDEAGTEFENYPOREKISERMLGHLTKGKEENGEPODYKEVH 131
Db      55  --GSP-----YDENLDD-----VEKATGQSTMAHL--NKSQDVG-----IHA 90
Qy      132  NMLPRLTSKODTSGEESAASPERLVSSTIAGKRLPISSDVNQSPNRRIYDPVGLGNA 191
Db      91  RHISVSTLIDSEMAEDN-----GNSKI 111
Qy      192  KKERVDGMKMKQOKNTGVPYSTQASERGVDDIDASTDILADEALLND---EAROPLSK 247
Db      112  KKNRVESMKERKKKKKPPATTKVERAE-----IPPOQMEDKRAPDASOPLSTY 161
Qy      248  VSISSRINPYRYAVIMLRVILCLFLHYRITNPVPNAPALMLVSVCEIMFALSWILDOF 307
Db      162  IPIPKRLAPYKRVIIIMRLIIGLFHYRYVTNPVDSAFGLMTLSVCEIMFALSWILDOF 221
Qy      308  PKWEPVNRRETYLDRLALRYDRGEPSQLAAVDIPEVTPLKPEPLVANTYLSTIADY 367
Db      222  PKWEPVNRRETYLDRLALRYDRGEPSQLAAVDIPEVTPLKPEPLVANTYLSTIADY 281
Qy      368  PVDKVCYVFDGAAMLSESLAETSEFARKWPFCKKYSIEPRAEWFAAKIDYLDKX 427
Db      282  PVDKVCYVFDGAAMLSESLAETSEFARKWPFCKKYSIEPRAEWFAAKIDYLDKX 341
Qy      428  VQTSFYKDRAMKREYEERKIRINALYSALKCPDEGWMODGTPWPGNNTGDHPGMIOY 487
Db      342  VQTSFYKDRAMKREYEERKIRINALYSALKCPDEGWMODGTPWPGNNTGDHPGMIOY 401
Qy      488  FLQONGDLAEENELPRLYVSREKRPQOHKKAGAMALVAVSVLNGPEPIINLDD 547
Db      402  FLQONGDLAEENELPRLYVSREKRPQOHKKAGAMALVAVSVLNGPEPIINLDD 461
Qy      548  HYNNKALREAMCFPLMDPQVGRDYCYVOPQRFDDIDKNDRYANRYTFDDIINLGLDG 607
Db      462  HYNNKALREAMCFPLMDPQVGRDYCYVOPQRFDDIDKNDRYANRYTFDDIINLGLDG 521
Qy      608  IGGPVYVGTGVFNRTALYGEPPKVKHKPDSLKLCGSRKKSKAKKESDKKSGR 667
Db      522  IGGPVYVGTGVFNRTALYGEPPKVKHKPDSLKLCGSRKKSKAKKESDKKSGR 579
Qy      668  HDSITVPVNRNDIEGVEGAGFDD--EKALIMSOMSLERKQSANVAFVASTLEMNGVPP 726
Db      580  REELDAALINLREID-----NDEYERKMLISQTSFEFTGLSVFESTLEMNGVAV 633
Qy      727  SATPENLLEALHIVISCGEYEDKSDMGMEIGWYIGSTEDILGFKMHARGMSIYCMPL 786
Db      634  SATPENLLEALHIVISCGEYEDKSDMGMEIGWYIGSTEDILGFKMHARGMSIYCMPL 693
Qy      787  PAFKGSAPINLSDRLOVLRMALGSVEILSRHCPITWY--NGRLKELERFAVNTTYIP 845
Db      694  PAFKGSAPINLSDRLOVLRMALGSVEILSRHCPITWY--NGRLKELERFAVNTTYIP 753
Qy      846  ITSPLIAMCTLAVCLFTNOFIIPOTSINASIIMPLSLFSTFATGILMRMSGVGIDEM 905
Db      754  ITSPLIAMCTLAVCLFTNOFIIPOTSINASIIMPLSLFSTFATGILMRMSGVGIDEM 813
Qy      906  WRNEQFVWJGVSAHLFAVFOGILKVLAGIDTNETYTSKASDEGDFALYLFKWTLLI 965
Db      814  WRNEQFVWJGVSAHLFAVFOGILKVLAGIDTNETYTSKASDEGDFALYLFKWTLLI 872

```





Db 181 PQRFEDGIDRHRYANRNRYFEDINMLGLDGLGPPYVGTGCVENRQALYGDPPVSEKR 240  
QY 637 KR-----PSILSKLGGSRKRSKRAKE-----SDKKSGRH--TDSTVPFNL 678  
Db 241 PKMTCCWPCWCCGCCGSRKSKKKEKGLLGLLGGKMKMKNNYKKSASAVFDL 300  
QY 679 DDIEGEGAGFDD--EKALLMSQMSLEKRGOSAVFVASTLMENGVPSPAPPENLLKEA 737  
Db 301 EEIEEGLE--GYEELKSTLMSQKNEKRRGQSPVFIASFLMENGGLPEGTNSTSLKEA 358  
QY 738 IHVSGYEDKSDMGMEIGMIVGYSVEDLITGFKMARGRSRYCPCPKLPAFGSAPINL 797  
Db 359 IHVSGYEKTEKMEIGMIVGYSVEDLITGFKMKRGKWSYCVKPKPAFGSAPINL 418  
QY 798 SDRNLQVLRWALGVSFLESRHCPYNGYNGRLKFEREAYVNTTYPIPSILWACTL 857  
Db 419 SDRNLQVLRWALGVSFLESRHCPYNGYNGRLKFEREAYVNTTYPIPSILWACTL 478  
QY 858 LAYCFTNPIIPQISINIASIMWLSLPSIFATGILEMWSGVGIDEMWREQFWVIGV 917  
Db 479 PAVCLLTGKFIITPTLSNLSVWFLALFLSIATGVELRMSGVSIQDMWNRQFWVIGV 538  
QY 918 SAHLFAVFGIILKVLADIDTNEFYNSKASDEGDAFELYFKTTLITPTLLIYNMG 977  
Db 539 SAHLFAVFGIILKVLADIDTNEFYNSKASDEGDAFELYFKTTLITPTLLIYNMG 597  
QY 978 WVAGSYAINSGYOSGMPPLSKLFFAFWYIVHLHYPLKGLMGRQNTPTIVVWSVLAS 1037  
Db 598 WVAGSDAINNCGSMGSPPLSKLFFAFWYIVHLHYPLKGLMGRQNTPTIVVWSVLAS 657  
QY 1038 IESLWVRIDPTSRVTGPDIEEGINC 1065  
Db 658 IESLWVRIDPTSRVTGPDIEEGINC 685

RESULT 11  
Db 86157  
hypothetical protein F22D16.26 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: D86157  
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, D.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: D86157  
A:status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1181 <STO>  
A:Cross-references: GB:AEO05172; NID:96056428; PIDN:AAF02892.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 40.98; Score 2324; DB 2; Length 1181;  
Best Local Similarity 44.0%; Pred. No. 2,4e-175;  
Matches 491; Conservative 165; Mismatches 268; Indels 192; Gaps 29;

QY 18 QTCQI--CSDNGKTYDGRFVADICSPVCRPCYEYERKDNQSCPOCKTKYKRLKS 75  
Db 160 QTCMLGCDG---KVVHG---RCE--CGFRICDCTFCDDTSGGCGPCGKEPYRIND 211  
QY 76 PAIPGCDGDLADEGTEVFENYPOKEKISERMLGMLHTRGKGEEMGEPOYDKESHNL 135  
Db 212 PEEDEDEDEAKP-----LPQ-----MESKIDRLKS---VY 241  
QY 136 RLTSRODTSGEFSASPERLSVSTIAGKRLPLSSDVNOSPNRRIVDPGLGNVAMKER 195

Db 242 KSEKAOQAGDEPHT-----RMLETK-----GTYGYNAAVPR 275  
QY 196 VQGMKKKQKKNPVSTQASERGVADIDSLADEALNDEAKOPLSRKYSIPSR 255  
Db 276 -DGYI-----GSGGGNGYEPPE-----FGRSRKPLTRKYSAAII 314  
QY 256 NPYRATIMRLVILCELYRTNPVPAFALMVSVICEIMFALSMIIDOPKPPVNR 315  
Db 315 SPYRLIALRVALGFLTWVRVHRPRAEMMLMGSTGCELMFALSMIIDOLPKLCPVR 374  
QY 316 EYLDRLALRYDEG--EP--SOLAVIDEYSTVDPLKEPLRYANTYLSIADY 370  
Db 375 LFDGLVKERFESPNLRNPKGRSDLDGIVFVSTADPEKPEPLVYANTYLSIADY 434  
QY 371 KVSQYEDDGAALSPESLAETSEFARKWVPCFKKYSIRPEWFAKIDPKKYOT 430  
Db 435 KLAGYSDGGALLTEALAOIASFASVWPCCRKINIEPRNDEAYFGOKRNLKAKYL 494  
QY 431 SPYKDRAMKREYEERKINAL-----VSKALC----- 460  
Db 495 DFYERRRYKREYDEKVRINSLPEAIRRSDAYVHLELRAKKKQEMMGNNGPQETVI 554  
QY 461 -PEEGVWODGTPWPG-----NTGHPGMIOVFLQNG-----GLDABGN----- 500  
Db 555 VPKATV-MSDGSHPGQTSGETNSRGDAGIIOMLAPNAEPYGAADAEMLDIT 613  
QY 501 ----ELPRIVYYSREKRGFOHKKAGANMALRVASVLTNGPFILNLCDDHYINSKAL 556  
Db 614 DVDIRLPMLYYSREKRGFOHKKAGANMALRVASVLTNGPFILNLCDDHYINSKAL 673  
QY 557 REAMCEFLMDNLQKQYVQFQRFQDIDKNDRYANRNTVFEDINLGLDGLGPPYVGT 616  
Db 674 BEGMCFLDRG--GDRICYVQFPQREFGIDPNDRYANHNTVFQVSRALDGLGPPYVGT 732  
QY 617 GCYFNRTALXGYEPLTKVHKKPSLSKLCGSRKRSKRAKESDKKSGRHTDVTVP 676  
Db 733 GCIFRRTALXGYEPPRATHH-----GWLGRKRVYISLRPRAMKKKDEVSPLIN 783  
QY 677 NLDIDIEGEGAFDEKALMSQMSLEKRGOSAVFVASTLMENG-- 722  
Db 784 GEYNEENDG-----DIESEL-----LPRFGNSNFVASIPVAEYQGLRIDDLQKGN 834  
QY 723 -----GVPSATPENLKEALHIVISCYEDKSDMGMEIGMIVGYSVEDLITGFKMAR 775  
Db 835 SRPAGSLAVRBDPLDAAVAEALSVISCYEDKTEGKRVGIVGSDVATVGYRMNR 894  
QY 776 GMRSTYCMKRLAFKCSAPINLSDRLNQLVRLMAGSVLELPSRHCIYWGYNRLKFLER 835  
Db 895 GMRSTYCMKRLAFKCSAPINLSDRLNQLVRLMAGSVLELPSRHCIYWGYNRLKFLER 952  
QY 836 FAYVNTTYPIPTISPLMCTLLAVCLFTNPFIPDISNIASIFLSFATGILM 895  
Db 953 VAFVNGMTPFISPLIYVCIPLAISLFGQPIVQSLDITLLIYLSITLITLCLMSL 1012  
QY 896 RMSGVGIDEMWREQFWVIGVSAHLFAVFGIILKVLADIDTNEFYNSKAS--DEGDD--F 952  
Db 1013 KMSGITLHEMWRNEQFWVIGVSAHLFAVFGIILKVLADIDTNEFYNSKAS--DEGDD--F 1072  
QY 953 AEIYLFKMTTLITPTLLIYNLVGVYVAINSGYOSGMPPLSKLFFAFWYIVHL 1012  
Db 1073 ADLYVWKSFLWVPLTIMVNVNIAVAGIARLVSPFOWSKLVGVFESFWVLCHLXP 1132  
QY 1013 FLKGLMGRQNTPTIVVWSVLASLPSIFATGILEMWSGVGIDEMWREQFWVIGV 1048  
Db 1133 FAKGLMGRQNTPTIVVWSVLASLPSIFATGILEMWSGVGIDEMWREQFWVIGV 1168

RESULT 12  
T51546  
cellulose synthase catalytic subunit-like protein - Arabidopsis thaliana  
N:Alternate names: protein F2K13\_60  
C:Species: Arabidopsis thaliana (mouse-ear cress)



C:Date: 18-Aug-2000 #sequence\_rev18on 18-Aug-2000 #text\_change 18-Aug-2000  
C:Accession: T51546  
R:Sato, S.; Nakamura, Y.; Kaneo, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mei  
submitted to the Protein Sequence Database, August 2000  
A:Reference number: 225394  
A:Accession: T51546  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1145 <SNT>  
A:Cross-references: EMBL:AL391141  
C:Genetics:  
A:Experimental source: cultivar Columbia; BAC clone F2K13  
A:Map position: 5  
A:Introns: 297/2; 566/3  
A:Note: F2K13\_60

Query Match 40.6%, Score 2303.5; DB 2; Length 1145;  
Best Local Similarity 43.8%; Pred. No. 9,66-174;  
Matches 490; Conservative 162; Mismatches 265; Indels 201; Gaps 26;

QY 19 TCQI--CSDNVGKTVGDGFVADICSEFPVGRCPYEXYERKDGNOGSPCKTRYRLKSGP 76  
DB 127 SCAIPGCDAKWMSDERGQDLTPCE--CDFKICRDFIDAVKTKGGICPCKBPYK---NT 181  
QY 77 AIPDDKDEDGLADGTYEFNFYPOKEKISERLGHHLTRGKEEMGEPEYDKEVSHNLPR 136  
DB 182 HLTDOVENG-----QQRPM---LPQGGGSKM----- 205  
QY 137 LTRQDHSGEFSAASPERLS--VSSTIAGKRFLPSSDVNOGPNRIVDPV---GLGNVAV 192  
DB 206 -----ERRLSMVKSTISALMSRSGTGDH--NMKLPEITTYTGYNAPW 248  
QY 193 KERVDGWMKQKNTGPVSTQAASERGVDIDASTDILADEAL--LNDEAROLSRKYSIP 251  
DB 249 -----TKDGFSGCKDGDGDDGDMGEADOLMSRPMPLRLKILP 289  
QY 252 SSRINPRAMVIMLVILCLELHRTINPVPVNAFALMVSYICETWFLSLIDQPKWF 311  
DB 290 AGVSPYRKLIFIRIVYLALELTWRVKHQNDAAVLMGMSYCELMFALSWILDQLPKLC 349  
QY 312 PVNRETYLDRLALRYD--REGEF--SOLAADVIFSVTDPLEPPLVTANTVLSILAVD 366  
DB 350 PINATDLOVLEKEFEPTPTASNPTGKSDLPGEFDVFVSTADEKEPEPLTANTVLSILAE 409  
QY 367 YPVNDKSCYVEDDGAAMLSEFSLAETSEFARKWVPFCCKYSIEBPAPWFAAKIDYLD 426  
DB 410 YPVEKLSGYVSDGCGALLTFEAMAEASFAIWPFCCKNAIEPRNDPSYSLKRDYKN 469  
QY 427 KVQVTSFYKDRAMKRETEEFKIRINALVS-----K 456  
DB 470 KVKSDPFYKDRRRVREDEFKRVNLSLPSDIRRSSDAYHAREETKAMKMQRONDDPEME 529  
QY 457 ALKPRPEGWVWQDSTPMPG-----NNTGDHGMTOVFLG-----QNGGLDAE 498  
DB 530 PVKIFPKATW--MADGTHMPGTLVTSASDHAKGDHAGIIQVMLKPPSDEPLHGVSSEGLDT 588  
QY 499 GNE--LEPLVYVSREKRFQHHKAGAMNLAIVSVASVLTNGPEPLINDCHYINNSKAL 556  
DB 589 DVDIRPLPLVYVSREKRGYDHNNKAGAMNLAIVASAIMSGPFLINDCHYIYNSAL 648  
QY 557 REAMCFLMDPLGKQVQCYVQRFQFDGIDKKDRYANRNTVFEDINLRGLDGIQGPVYGT 616  
DB 649 REGCCPFMWDRC--GDRLCYVQFPQRFEGIDPDRANHTVTFEDVNMRLADGLMPPVYGT 707  
QY 617 GCVENRTALVYEPPIVYKHKPSLSLKGSGSRKN-----SKAKESDKKKSGRHDS 672  
DB 708 GCLRRRLALVYFNPP--RSKDFSPSCWSCCPRRSRKKIPENRRLRMSD----- 755  
QY 673 VPVENLDDIEBGVAGGAFDEKALIMSQMLEKRFQSGAVAS-----TL 718  
DB 756 -----YDDEE---MNLSTVPKFKFGNSTFLIDSLPAVAEPGGRPLADHPA 795  
QY 719 MENGVSPPSAT--PENLL-----KRAIIVISGCTEDKSDMGMEIGMIGSYTEDILITGFM 772

796 VKNGRPPALLIIPRELLDASVYAEIAIVTSCHTEDTEGSRIGMAYIGSTEDVGYGRM 855  
 QY 773 HARGRSITYCPKLEPAFKGSAPINLSDRLNOVLRMALGSEYIELSRHCPIWYGNRLTK 832  
 Db 856 HNRGKSVSYCYTKRPAFGTAPINLDRHQVLRMATGSEVEIFFSNMNL--ASSKMKI 913  
 QY 833 LERRAYVNTTYPTISIDPLMCTLANCLFNNQFIIPDISNATSWFLSLFISTPATGI 892  
 Db 914 LQRIAYLVNGVYIPFETISFLIVCYFLPALSLFSGOEFYQTLNWTFLVLLIISTLCIAL 973  
 QY 893 LEMRSGCIDEMWNEQFVWIGVSAHLFAVFOGLTKYLAGIDTFWTYSKA--SDEG 950  
 Db 974 LEIMWSGISLEEMWNEQFWLIGTSAHLAAVQGLKAVAGVEISFTLSKSGDDID 1033  
 QY 951 DPAELYLKMTTLIPPTLLILVNLGVYAGVSAVYAINSGVSGPFLGKLPFAFWIVHL 1010  
 Db 1034 EFADLYMWKMSLMIRPTTIIMVNLIAVGSRTIYSVVPQMSKILGIVFSEFWLAHL 1093  
 QY 1011 YPFLKGLMGONRPTIYVWMSVLLASIFSLMWRIDP 1048  
 Db 1094 YPFAKGLMGRGRPTIYVWMSGLVAITSLMWAINP 1131  
 RESULT 13  
 T05646  
 hypothetical protein F20D10.310 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
 C:Accession: T05646  
 R:Byan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer  
 submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15420  
 A:Accession: T05646  
 A:Molecule type: DNA  
 A:Residues: 11111 <BEV>  
 A:Cross-references: EMBL:AL035538  
 A:Experimental source: cultivar Columbia; BAC clone F20D10  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 139/22: 675/3  
 A:Note: F20D10.310  
 Query Match 40.3%; Score 2289.5; DB 2; Length 1111;  
 Best Local Similarity 43.1%; Pred. No. 1.2e-172;  
 Matches 481; Conservative 158; Mismatches 256; Indels 221; Gaps 28;  
 QY 23 CSDNVGKTVNDGRFVACDTCSPVCRPCYEYERKDNOSGPOCKTRYKRLKSGPAIPGDK 82  
 Db 120 CDGNVMKDERGKDVMPCE-CRFKICRDCPMDAOKE-TGLCPGCKEYK-----IGDL 169  
 QY 83 DEDGL-ADEGTVENYPOKEKISEKMLGHLTRGKEEGEPQYDKVEVSHNHLPRLSRQ 141  
 Db 170 DDDPDPVYSSGALPLPAPGKDO-----RGNNNNS-----MKR 202  
 QY 142 DTSGEFSAASPERLSVSSSTIAGKRLPYSSDVNQSPPRRIVD--DVGLGNVAM-KERYD 197  
 Db 203 NQNGEF-----DHNRLMEFYEGYGVGNATWMPQDEM 234  
 QY 198 GWAKQKQENGPVSTQAASERGGVIDASTDLADEALLDEAKRQPLSRKVSIPSSRINP 257  
 Db 235 GDDMDEGRGMVETA-----DKPWRPLSRIRIPPAIISP 270  
 QY 258 YRMVIMLVLVLTLCLEFIRITNPVNAFALMLVSVICETIMPALMSWILDDPPKFPVNRRT 317  
 Db 271 YRLIVLIVFVYVLCFLWLRINPNEDATWMLMSTICELMFGESWILDDQPKLCPINRST 330  
 QY 318 YLDRLALRYDEGEDE-----SOLAANDIFVSTVDLPKEPPLVANTVLSILADVDPKV 372  
 Db 331 DLEVLRDKFDMPSPENPTGRSDLPGLDVFVSTADPEKPPVLVTANTILSLIANDVYEKV 390  
 QY 373 SCYVFDGAMLSFSLAETSFAKRWKVPCKKYSIEPAPEMYPFAKIDLYLKKQYOTSF 432  
 Db 432

Db 391 SCVLSDDGALLSTEAMAEASPADLWVPCRKHNIEPRNDSTFSLKIDPTKKSIDF 450  
 QY 433 VKDRRAKREYEEKIRINALVS-----KAL-----KCPRE 463  
 Db 451 VKDRRAKREYEEKIRINALVS-----KAL-----KCPRE 463  
 QY 464 GWMVODCTPPRG-----NNTGDHGMVQVFLGQNGGDAEGN-----EL 502  
 Db 511 TM-MADGTHWPGWAASTREHSGKDAGILLQVLMKPPSSDPLIGNSDKYIDFSDTDTL 569  
 QY 503 PRLVYRSREKRGFOHKKAGAMALVRSVAVLTNGPFILNDCDHYINNSKALREAF 562  
 Db 570 PRLVYRSREKRGFOHKKAGAMALVRSVAVLTNGPFILNDCDHYINNSKALREAF 562  
 QY 563 LMDPNLQYQYVQFQDFQIDKNDRYANRNTVFEDINLGLDQVYVGVGTCVFN 622  
 Db 630 MMDRG-CEGICVIOFQFQDFQIDKNDRYANRNTVFEDINLGLDQVYVGVGTCVFN 622  
 QY 623 TALXGYPPIKVKHKKSLSKLGGSKRKNKAKKSDKKSGRHTDSTVPVFNLDIE 682  
 Db 689 FALYGEPP-----NPKLE-----KKESE-----T 710  
 QY 683 EGVEGAGFDEKALLMSOMSLERKFGOS-----AVFVASTLMNGV---PESA- 728  
 Db 711 EALTSPEDP-----LDYQLPKRFNGSTLAEISIPAEQGRPLADHPAKYGRPGAL 766  
 QY 729 -TRENLL-----KEAHVISCYEDKSDMGMEIGVSYTEDILNCFKMHARGRSIYC 782  
 Db 767 RVRDPLDATTVAESVSICWYEDKEMGDVGMISYTEDVYVGRHNGMRSYVC 826  
 QY 783 MRLPAKGSAPINLSRLNOVRLMALGSVEILFSRCPITWYNGKLEKLEPVAVNTT 842  
 Db 827 ITRDSFRSGAPINLSRLNOVRLMALGSVEILFSRCPITWYNGKLEKLEPVAVNTT 842  
 QY 843 IYPTISLPIMLCTLLAVCLFTNOFTIPOSINIASIEMSLFSLFSTFGLIEMKWSGI 902  
 Db 885 IYPTISLPIMLCTLLAVCLFTNOFTIPOSINIASIEMSLFSLFSTFGLIEMKWSGI 902  
 QY 903 DEWMRNEQFVWIGVSAHLEFVQGLIKVLAGIDTNTVTSKASDEGD--FAELYLFKW 960  
 Db 945 EEWWRNEQFVWIGVSAHLEFVQGLIKVLAGIDTNTVTSKASDEGD--FAELYLFKW 960  
 QY 961 TPTLLIPTLLIYNVGVAVGSAVNSGQSGPLFGKLFPAFVYVHLVPLKGLMGR 1020  
 Db 1005 SSIMIPPIVAMNIIAIVAFIRITVQAVPOMSLGAFEFVWVLAHLVPLKGLMGR 1064  
 QY 1021 QNRTPTIVVSVLASISFLMWRIIDPFTSVTGP 1056  
 Db 1065 RGTPTIVVSVLASISFLMWRIIDPFTSVTGP 1056  
 RESULT 14  
 D84741  
 C:Species: Arabidopsis thaliana (mouse ear cross)  
 C:Date: 02-Feb-2001 #sequence: revision 02-Feb-2001 #text: change 02-Feb-2001  
 C:Accession: D84741  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: D84741  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1036 <STO>  
 A:Cross-references: GB:AE002093; NID:92924781; PIDN:AA04910.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: Atg331100  
 A:Map position: 2  
 Query Match 39.1%; Score 2221.5; DB 2; Length 1036;

Best Local Similarity 45.0%; Pred. No. 2,6e-167;  
 Matches 469; Conservative 150; Mismatches 248; Indels 175; Gaps 26;  
 QY 117 GEEMGEPOYKEVSHNLPRLTSRODTSGEFSAASPRLSVSTIAG-----KR 166  
 Db 48 GDYSGNDY---INTYLMPTPDNCPAGSSGTSSEKGANCGGGDDPKMKNKLE 104  
 QY 167 LPSYSDNQS-----PNRIVD---PYGLGVAMKERVYDKMKOEKNTGCVST 212  
 Db 105 LSVKSNKSMILRSQGDGDFDHNRLVESGKYGIGNAFSEEDDTVD----- 152  
 QY 213 QASERGVQIDASTDLADLADLNDARPLRSKVSIPSSRNIPRYMIVMLVILCLF 272  
 Db 153 -GGV---SKSDPL-----DKPMKPLTRKQVDPKILSPYRLVILVYVFF 196  
 QY 273 LHVITPVPNAFALMVSVICETWPAFSLTLOFPWFVFNRETYLDRLALRDRGEP 332  
 Db 197 LHMRIITPNEDAMLMGLSTVCEIWFASFILDLPLNPNAT---DLAALHDFEOP 253  
 QY 333 -SOLAADVIFVSTVDPLEKPPVTANTVLSIADVVPVDSVSCVFDGGAAML 384  
 Db 254 SPSPNTRSDLPQVDVFSVTADEKEPPLVTANTLILAVDPIELKSAVISDDGAIL 313  
 QY 385 SPESLATSEFAKKVWPCKKYSIEPPAPWYTAADYLDKDVQVSVFVDRAMKREYE 444  
 Db 314 TFEAMAEVAFVAFVWPCFKHIEPRNPDSYSEIKDPYKKNKRODFVDRMIREYD 373  
 QY 445 EFRIRIALYSKALK-----CPREG-----VWVODCTPW 473  
 Db 374 EFKYRINGLPEQIKKRAEOPNMEELKEKRIAEKNGVLPFGVEVYKATW-MADGTHW 432  
 QY 474 PG-----NNTGDHGMVQVFL-----GON-GGLDAEGNEL--PRLVYRSREK 512  
 Db 433 PGTFEPEKPDHSGKDAGIILQIMSKVPDLEPVMGPRNEGALDFTGIDIVPMFAVYRSREK 492  
 QY 513 RRGFOHKKAGAMALVRSVAVLTNGPFILNDCDHYINNSKALREAFMCLMPNIGKOV 572  
 Db 493 RRGFOHKKAGAMALVRSVAVLTNGPFILNDCDHYINNSKALREAFMCLMPNIGKOV 572  
 QY 573 CYVQFQDFQIDKNDRYANRNTVFEDINLGLDQVYVGVGTCVFNFTALYGEPP 632  
 Db 552 CYVQFQDFQIDKNDRYANRNTVFEDINLGLDQVYVGVGTCVFNFTALYGEPP 632  
 QY 632 KVKHKKRSLSKLGGSKRKNKAKKESDKKSGRHTDSTVPVFNLDIEEGVAGFDD 692  
 Db 611 -RANEXSGVFGQEKAPAMHVRTQSQASQTSQ-----ASDLESPTQPLNDP 655  
 QY 655 EKALMSQMSLERKFGOSANF-----VASTLMNGCVPSA-----TPEN 732  
 Db 656 D-----LGLPKRFNGSTFETDTIPVAEYQGRPLADHMSVKNKGRPGALLRPPPLDAP 708  
 QY 732 LKKEAHVISCYEDKSDMGMEIGVSYTEDILNCFKMHARGRSIYCMPLPAFKGS 792  
 Db 709 TVAEALAVISCWEDNTEMKDRIGVYSTEDVYVGRHNGMRSYVCITRDAFRGT 768  
 QY 793 APINLSRLNOVRLMALGSVEILFSRCPITWYNGKLEKLEPVAVNTTIVYTSIPL 852  
 Db 769 APINLSRLNOVRLMALGSVEILFSRCPITWYNGKLEKLEPVAVNTTIVYTSIPL 852  
 QY 852 MYCTLLAVCLFTNOFTIPOSINIASIEMSLFSLFSTFGLIEMKWSGI 902  
 Db 827 VYCFLEPLCLFSKFTVOSL---DIHPSYLLCTIVTTLTSLLEKWSGILEEMWRN 882  
 QY 909 EGFVIGVSAHLEFVQGLIKVLAGIDTNTVTSKASDEGD--FAELYLFKW 966  
 Db 883 EGFVIGVSAHLEFVQGLIKVLAGIDTNTVTSKASDEGD--FAELYLFKW 966  
 QY 967 PTTLLIYNVGVAVGSAVNSGQSGPLFGKLFPAFVYVHLVPLKGLMGRNRTPT 1026  
 Db 943 PTTLLIYNVGVAVGSAVNSGQSGPLFGKLFPAFVYVHLVPLKGLMGRNRTPT 1026  
 QY 1027 IVVYVSVLASISFLMWRIIDP 1048



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:08:04 ; Search time 32 Seconds  
(without alignments)  
3558.288 Million cell updates/sec

Title: AAC393336

Perfect score: 5677

Sequence: 1 MESEGETAGKPKMKNIPTQC.....IDPFTRVATGPDILEGIMC 1065

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5677	100.0	1065	US-09-900-237-33	Sequence 33, Appl
2	4507	79.4	1080	US-09-900-237-30	Sequence 30, Appl
3	4133.5	72.8	881	US-09-838-539-8	Sequence 8, Appl
4	3999.5	70.5	1091	US-09-900-237-26	Sequence 26, Appl
5	3922	69.1	1148	US-09-900-237-4	Sequence 4, Appl
6	3873.5	68.3	1086	US-09-900-237-10	Sequence 10, Appl
7	3843.5	67.7	1165	US-09-900-237-8	Sequence 8, Appl
8	3804.5	67.0	1039	US-09-900-237-14	Sequence 14, Appl
9	3483.5	61.4	701	US-09-900-237-32	Sequence 32, Appl
10	3473.5	61.2	974	US-09-838-539-6	Sequence 6, Appl
11	3216	56.6	793	US-09-900-237-18	Sequence 18, Appl
12	3003.5	52.9	740	US-09-900-237-24	Sequence 24, Appl
13	2808	49.5	685	US-09-900-237-31	Sequence 31, Appl
14	2803	49.4	685	US-09-838-539-7	Sequence 7, Appl
15	2647.5	46.6	610	US-09-900-237-16	Sequence 16, Appl
16	2420	42.6	506	US-09-900-237-20	Sequence 20, Appl
17	2055.5	36.2	431	US-09-900-237-28	Sequence 28, Appl
18	1800	31.7	720	US-10-260-046-18	Sequence 18, Appl
19	1405	24.7	522	US-10-260-046-2	Sequence 2, Appl

20	1405	24.7	522	9	US-10-260-046-4	Sequence 4, Appl
21	1388.5	24.5	320	10	US-09-900-237-6	Sequence 6, Appl
22	1358.5	23.9	304	10	US-09-900-237-2	Sequence 2, Appl
23	1152.5	20.3	340	10	US-09-900-237-22	Sequence 22, Appl
24	1071	18.9	590	9	US-10-260-046-30	Sequence 30, Appl
25	1044	18.4	727	9	US-10-260-046-24	Sequence 24, Appl
26	1015	17.9	741	9	US-10-260-046-28	Sequence 28, Appl
27	594	10.5	341	10	US-09-900-237-12	Sequence 12, Appl
28	505.5	8.9	166	10	US-09-734-566-154	Sequence 154, App
29	499.5	8.8	159	10	US-09-734-566-40	Sequence 40, Appl
30	498.5	8.8	165	10	US-09-734-566-38	Sequence 38, Appl
31	289	5.1	115	10	US-09-734-566-42	Sequence 42, Appl
32	278	4.9	693	10	US-09-838-539-11	Sequence 11, Appl
33	245.5	4.3	861	10	US-09-838-539-12	Sequence 12, Appl
34	236.5	4.2	756	10	US-09-838-539-10	Sequence 10, Appl
35	196	3.5	723	10	US-09-838-539-9	Sequence 9, Appl
36	118.5	2.1	1177	9	US-10-024-623-20	Sequence 20, Appl
37	118.5	2.1	1177	9	US-09-795-927-2	Sequence 2, Appl
38	115	2.0	357	10	US-09-874-133-27	Sequence 27, Appl
39	107.5	1.9	1221	9	US-10-270-333-60	Sequence 60, Appl
40	107.5	1.9	1325	9	US-10-154-452-6	Sequence 6, Appl
41	107.5	1.9	1437	9	US-10-154-452-2	Sequence 2, Appl
42	107.5	1.9	1437	9	US-09-934-421A-6	Sequence 6, Appl
43	107.5	1.9	1437	9	US-10-162-012-38	Sequence 38, Appl
44	106	1.9	357	10	US-09-874-133-26	Sequence 26, Appl
45	104.5	1.8	910	9	US-10-012-140-17	Sequence 17, Appl

#### ALIGNMENTS

RESULT 1  
US-09-900-237-33  
Sequence 33, Application US/09900237  
Patent No. US20020120124A1  
GENERAL INFORMATION:  
APPLICANT: Allen, Stephen  
TITLE OF INVENTION: Plant Cellulose Synthases  
FILE REFERENCE: B81170 US CIP  
CURRENT APPLICATION NUMBER: US/09/900,237  
CURRENT FILING DATE: 2001-07-06  
PRIOR APPLICATION NUMBER: 60/092,844  
PRIOR FILING DATE: 1998-07-14  
PRIOR APPLICATION NUMBER: PCT/US99/15871  
PRIOR FILING DATE: 1999-07-13  
PRIOR APPLICATION NUMBER: 09/720383  
PRIOR FILING DATE: 2000-12-21  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 33  
LENGTH: 1065  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-09-900-237-33  
Query Match 100.0%; Score 5677; DB 10; Length 1065;  
Best Local Similarity 100.0%; Pred No. 0;  
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MESEGETAGKPKMKNIPTQCISDNVGTVDRETVACDICSFPVCRPCYEYERKDGNO 60  
|||||  
1 MESEGETAGKPKMKNIPTQCISDNVGTVDRETVACDICSFPVCRPCYEYERKDGNO 60  
Db 1 MESEGETAGKPKMKNIPTQCISDNVGTVDRETVACDICSFPVCRPCYEYERKDGNO 60  
QY 61 SCPOCKTRKRLKLGSAITGDKDDEGLABEGTYEFNYPQKEKISERMMLHMLTRGGEEM 120  
|||||  
61 SCPOCKTRKRLKLGSAITGDKDDEGLABEGTYEFNYPQKEKISERMMLHMLTRGGEEM 120  
Db 61 SCPOCKTRKRLKLGSAITGDKDDEGLABEGTYEFNYPQKEKISERMMLHMLTRGGEEM 120  
QY 121 GEPOYKEVSHNLPPLTSGROPTSGEFSASPERLSVSSITAGGKRLPSSDVNOSPNNR 180  
|||||  
121 GEPOYKEVSHNLPPLTSGROPTSGEFSASPERLSVSSITAGGKRLPSSDVNOSPNNR 180  
Db 121 GEPOYKEVSHNLPPLTSGROPTSGEFSASPERLSVSSITAGGKRLPSSDVNOSPNNR 180  
QY 181 YDVPGLGVANKERVYDGKMKRQEKNTGPGVSTQOASERGVNDIDASTDILADEALINDEA 240  
|||||  
181 YDVPGLGVANKERVYDGKMKRQEKNTGPGVSTQOASERGVNDIDASTDILADEALINDEA 240

Db 181 IYDPEGLGNVAMKERVYDGMKMKOEKNTGPVSTQAASERGVDIDASTDILLADEALLNDEA 240  
QY 241 RQPLSRKVSIPSSSRINPYRMVIMLRVYLICLFHYRTITPNVPAFALMLVSVICELMFAL 300  
Db 241 RQPLSRKVSIPSSSRINPYRMVIMLRVYLICLFHYRTITPNVPAFALMLVSVICELMFAL 300  
QY 301 SWLLDOPPKWFPVNRRETYLDRALARYDREGESQLAVIDFVSTVDPLEKPEPLVTANTVL 360  
Db 301 SWLLDOPPKWFPVNRRETYLDRALARYDREGESQLAVIDFVSTVDPLEKPEPLVTANTVL 360  
QY 361 SIILAVYVPVAKVCYFDDGAAMLSPESLAETSEFARKWVPCKKXSTIEPRAPENYFAAK 420  
Db 361 SIILAVYVPVAKVCYFDDGAAMLSPESLAETSEFARKWVPCKKXSTIEPRAPENYFAAK 420  
QY 421 IDYLKRVQTSFVYKDRAMKREYEERKIRINALVSKALCPREGWYMODTTPMGNNITGD 480  
Db 421 IDYLKRVQTSFVYKDRAMKREYEERKIRINALVSKALCPREGWYMODTTPMGNNITGD 480  
QY 481 HPGMIOVFLGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSAYLTNGPF 540  
Db 481 HPGMIOVFLGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSAYLTNGPF 540  
QY 541 IILNDDCHYINNSKALREAMCFIADBNLCKQVCYVOPORPDGIDKNDRIANNTVPEDI 600  
Db 541 IILNDDCHYINNSKALREAMCFIADBNLCKQVCYVOPORPDGIDKNDRIANNTVPEDI 600  
QY 601 NLNGLDGLOGPVYVGTGVFNRTALYGEPPKIKVKKHKKPSLSKLCGSGSRKKSKAKKES 660  
Db 601 NLNGLDGLOGPVYVGTGVFNRTALYGEPPKIKVKKHKKPSLSKLCGSGSRKKSKAKKES 660  
QY 661 DKRSKRHTDSTVPVNNDDIEBGVAGAGFDDKALLMSQSLKREKRGOSAVVASTLME 720  
Db 661 DKRSKRHTDSTVPVNNDDIEBGVAGAGFDDKALLMSQSLKREKRGOSAVVASTLME 720  
QY 721 NGGVPSATPENILKRAIHVISCYEDKSPMGMEIGMIVGYSVEDIILTGKMARGRST 780  
Db 721 NGGVPSATPENILKRAIHVISCYEDKSPMGMEIGMIVGYSVEDIILTGKMARGRST 780  
QY 781 YCMLPAPKAGSAPINLSDRNLQVLRWALGSVEILFSRHCPIMYGVYNGRLKFLERRAYVN 840  
Db 781 YCMLPAPKAGSAPINLSDRNLQVLRWALGSVEILFSRHCPIMYGVYNGRLKFLERRAYVN 840  
QY 841 TTYPTTSTPLMAYCTLLAVCLFTNOFITPOISNIASIMPLSLFSLFAGILEMRSGV 900  
Db 841 TTYPTTSTPLMAYCTLLAVCLFTNOFITPOISNIASIMPLSLFSLFAGILEMRSGV 900  
QY 901 GIDEMMRNEDFWYIGVSAAHLFAVFOGILKVLAGIDTNTVTSKASDEDDFALYLFK 960  
Db 901 GIDEMMRNEDFWYIGVSAAHLFAVFOGILKVLAGIDTNTVTSKASDEDDFALYLFK 960  
QY 961 TTYLIPPTLLIVNLGVVAGVSYAINSGYSWGPGLKLFPAFWYIVHLYPELKGIMGR 1020  
Db 961 TTYLIPPTLLIVNLGVVAGVSYAINSGYSWGPGLKLFPAFWYIVHLYPELKGIMGR 1020  
QY 1021 QNRTPTIVVWSVLAISFSLMVRIDPFSRTYTGDIIECGINC 1065  
Db 1021 QNRTPTIVVWSVLAISFSLMVRIDPFSRTYTGDIIECGINC 1065  
RESULT 2  
US-09-900-237-30  
Sequence 30, Application US/09900237  
Patent No. US20020120124A1  
GENERAL INFORMATION:  
APPLICANT: Allen, Stephen  
TITLE OF INVENTION: Plant Cellulose Synthases  
FILE REFERENCE: BB1170 US CIP  
CURRENT APPLICATION NUMBER: US/09/900, 237  
CURRENT FILING DATE: 2001-07-06  
PRIOR APPLICATION NUMBER: 60/092, 844  
PRIOR FILING DATE: 1998-07-14  
PRIOR APPLICATION NUMBER: PCT/US99/15871  
PRIOR FILING DATE: 1999-07-13

QY PRIOR APPLICATION NUMBER: 09/720383  
QY PRIOR FILING DATE: 2000-12-21  
QY NUMBER OF SEQ ID NOS: 33  
QY SOFTWARE: Microsoft Office 97  
QY SEQ ID NO 30  
QY LENGTH: 1080  
QY TYPE: PRT  
QY ORGANISM: Trilicium aestivum  
QY US-09-900-237-30  
Query Match 79.4%; Score 4507; DB 10; Length 1080;  
Best Local Similarity 78.4%; Pred. No. 0;  
Matches 836; Conservative 98; Mismatches 107; Indels 26; Gaps 8;  
QY 20 CQICSNVGTVDGDFVACDICSFVCRPCYERERKDNQSCPOCKTRYKRLKSPAP 79  
Db 19 CQICADGLGTLDGVDFTACDVCRFVCRRCYEHKEKEGTQALQCKTKRRRSGPAIR 78  
QY 80 GDKDEDGLADEGTEVFENP-----OKERISRMLCMHLTRKGGEKMEGPQYDK----- 127  
Db 79 GEEGDDTDADDDGS-DENYPASGTEDKOKIADRRMRSMNMTGSGGVGHKPYDSGEIGLS 137  
QY 128 -----EYSHNHLPRLSRQDTSGEFSAASPERLSVSTIAGKRLPYSSDVNOSPRLIV 182  
Db 138 KIDSGETPQYVSVYVINSQ-MSGELPGASPDHMHMSPTGNISIRAPFPY-VNHSPPSPRE 195  
QY 183 DPGVGLGNVAMKERVYDGMKMKOEKNTGPV--STQAASE-RGVYDIDASTDILLADEALLND 238  
Db 196 FSGSIGNVAMKERVYDGMKMKOEKNTGPV--STQAASE-RGVYDIDASTDILLADEALLND 255  
QY 239 EAPQPLSRKVSIPSSSRINPYRMVIMLRVYLICLFHYRTITPNVPAFALMLVSVICELMF 298  
Db 256 ETPQPLSRKVSIPSSSRINPYRMVIMLRVYLICLFHYRTITPNVPAFALMLVSVICELMF 315  
QY 299 ALSMTIDOPPKWFPVNRRETYLDRALARYDREGESQLAVIDFVSTVDPLEKPEPLVTANT 358  
Db 316 ALSMTIDOPPKWFPVNRRETYLDRALARYDREGESQLAVIDFVSTVDPLEKPEPLVTANT 375  
QY 359 VLSILAVDVPVAKVCYFDDGAAMLSPESLAETSEFARKWVPCKKXSTIEPRAPENYFAAK 418  
Db 376 VLSILAVDVPVAKVCYFDDGAAMLSPESLAETSEFARKWVPCKKXSTIEPRAPENYFAAK 435  
QY 419 AKIDYTKDKVQTSFVYKDRAMKREYEERKIRINALVSKALCPREGWYMODTTPMGNNIT 478  
Db 436 QKIDYTKDKVQTSFVYKDRAMKREYEERKIRINALVSKALCPREGWYMODTTPMGNNIT 495  
QY 479 GDRHGMIOVFLGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSAYLTNG 538  
Db 496 RDRHGMIOVFLGQNGDLAEGNELPRLVYVSREKRPFGQHHKAGAMNALVRSAYLTNG 555  
QY 539 PFLNLDCHYINNSKALREAMCFIADBNLCKQVCYVOPORPDGIDKNDRIANNTVPEDI 598  
Db 556 QVNLNLDCHYINNSKALREAMCFIADBNLCKQVCYVOPORPDGIDKNDRIANNTVPEDI 615  
QY 599 DYNLRLDGIQGVYVGTGVFNRTALYGEPPKIKVKKHKKPSLSKLCGSGSRKKSKAKKES 658  
Db 616 DYNLRLDGIQGVYVGTGVFNRTALYGEPPKIKVKKHKKPSLSKLCGSGSRKKSKAKKES 673  
QY 659 ESDKRSKRHTDSTVPVNNDDIEBGVAGAGFDDKALLMSQSLKREKRGOSAVVASTLME 718  
Db 674 SSDKRSKRHTDSTVPVNNDDIEBGVAGAGFDDKALLMSQSLKREKRGOSAVVASTLME 733  
QY 719 MENGVPSPATPENILKRAIHVISCYEDKSPMGMEIGMIVGYSVEDIILTGKMARGR 778  
Db 734 MEYGVPOSTPESLILKRAIHVISCYEDKSPMGMEIGMIVGYSVEDIILTGKMARGR 793  
QY 779 STYCMKLPAPKAGSAPINLSDRNLQVLRWALGSVEILFSRHCPIMYGVYNGRLKFLERRAY 838  
Db 794 SYVCMKLPAPKAGSAPINLSDRNLQVLRWALGSVEILFSRHCPIMYGVYNGRLKFLERRAY 853  
QY 839 VNTTPTIVVWSVLAISFSLMVRIDPFSRTYTGDIIECGINC 898  
Db 854 INTTPTIVVWSVLAISFSLMVRIDPFSRTYTGDIIECGINC 913

Oy	899	GVGIDEMWNRNQFVAVIGVSAHLHFAVQIGILKVLVAGIDNTNTYTSKASDEGDAELTYL	958
Db	914	GVGIDEMWNRNQFVAVIGVSAHLHFAVQIGILKVLVAGIDNTNTYTSKASDEGDAELTYL	973
Oy	959	KWTLILPPTLLIVNLVGVAVGSAVYAINSGYSGMGLGALFPAFWYIVHLYPEFLGIM	1018
Db	974	KWTLILPPTLLIVNLVGVAVGSAVYAINSGYSGMGLGALFPAFWYIVHLYPEFLGIM	1033
Oy	1019	GRQNRPTIIVVWMSVLLASIFSLMWRIDPPTSRYVGPBIDLEGGINC	1065
Db	1034	GRQNRPTIIVVWMSVLLASIFSLMWRIDPPTSRYVGPBIDLEGGINC	1080
RESULT 3			
US-09-838-539-8			
; Sequence 8, Application US/09838539			
; Patent No. US20020129401A1			
; GENERAL INFORMATION:			
; APPLICANT: Stalker, D. et al.			
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter			
; FILE REFERENCE: 15621/03/US			
; CURRENT APPLICATION NUMBER: US/09/838,539			
; PRIOR FILING DATE: 2001-04-18			
; PRIOR APPLICATION NUMBER: 60/029,987			
; PRIOR FILING DATE: 1996-10-29			
; PRIOR APPLICATION NUMBER: 08/960,048			
; PRIOR FILING DATE: 1997-10-29			
; NUMBER OF SEQ ID NOS: 12			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 8			
; LENGTH: 881			
; TYPE: PRT			
; ORGANISM: Oryzae sativa			
US-09-838-539-8			
Query Match 72.8%; Score 4133.5; DB 10; Length 881;			
Best Local Similarity 86.8%; Pred. No. 0;			
Matches 766; Conservative 55; Mismatches 56; Indels 5; Gaps 3;			
Oy	188	GNVAMKEPVNDGKKKQKQKNTGPRV---STQASSEKGV-DIDASDILADEALLNDEARQP	243
Db	1	GNVAMKEPVNDGKKKQKQKNTGPRV---STQASSEKGV-DIDASDILADEALLNDEARQP	60
Oy	244	LSRKYSIPSSRINPYRMYIMLRVYLCLFHYRITNPVNAFALMVSGICEINFALSMI	303
Db	61	LSRKYPRLSSINPRNYIVLRVYLSFLHRYRITNPVNAFALMVSGICEINFALSMI	120
Oy	304	LDOFPKPEPVNRETYLDRLALRYDREGESQLAADVIFSVTDPLKEPRLVANTVLSIL	363
Db	121	IDQPKPMPINRETYLDRLALRYDREGESQLAADVIFSVTDPLKEPRLVANTVLSIL	180
Oy	364	AVDYPVDKVCYVDDGAGMLSEFSLATSEFAKKWPFCKKYSIEPAPEMYRAAKIDY	423
Db	181	AVDYPVDKVCYVDDGAGMLSEFSLATSEFAKKWPFCKKYSIEPAPEMYRAAKIDY	240
Oy	424	LKDYVQTSFVYDRBRAMKREYEEFRIKIALVSKALKCEBEGVMODGRPWPGNNTGDHG	483
Db	241	LKDYVQTSFVYDRBRAMKREYEEFRIKIALVSKALKCEBEGVMODGRPWPGNNTGDHG	300
Oy	484	MIQVFLQNGSLDAEGNELPRLVUYVSRKRPGRFOHNKKAGAMNLVYSAVLITNGPILN	543
Db	301	MIQVFLQNGSLDAEGNELPRLVUYVSRKRPGRFOHNKKAGAMNLVYSAVLITNGPILN	360
Oy	544	LDCHYIINNSALREAMCFLMDPNLKGQVYVORPQFQIDKNDRYANRNTVEFDINLR	603
Db	361	LDCHYIINNSALREAMCFLMDPNLKGQVYVORPQFQIDKNDRYANRNTVEFDINLR	420
Oy	604	GLDGIQGVUYVGTGCVFNRTALYGERPIYKHKKPSLLSLKCGSRKKNSKAKKESQK	663
Db	421	GLDGIQGVUYVGTGCVFNRTALYGERPIYKHKKPSLLSLKCGSRKKNSKAKKESQK	479

OY	664	KSGRHTSDTVAVENPNDIDEEVBEAGPDEKALLMSQSLKRGQSAFVFASTLMENGG	722
Db	480	KSNNHVDVAVFPNLEDIEEGVAGGFDEKSLMSQSLKRGQSAFVFASTLMETGG	539
OY	724	VPSATPEPNLKEAIHVISCYGEDKSDMGMEIGWYGSVTEEDILTGFFMHAHGRMSICYM	783
Db	540	VPSATPEPNLKEAIHVISCYGEDKTEWGTETLGMWYGSVTEEDILTGFFMHAHGRMSICYM	599
OY	784	PKLPAFKSGADPINSDRLNOVLRNALGSEVILFSRHCPYWTYGNRLKFLERFAYVNTTI	843
Db	600	PKRPAFKSGADPINSDRLNOVLRNALGSEVILFSRHCPYWTYGNRLKFLERFAYVNTTI	659
OY	844	YPLISIPILMCTLLAVGLFTNOFIPIPOISINASIWPLSLSTISFATILKMRMSGVIGID	903
Db	660	YPLISIPILYICVLPALICLLTGKFLIPEISINASTWPLSLSTISFATILKMRMSGVIGID	719
OY	904	EMWRNEQFWVIGVGSAAHLFAVFOGILKVLAGIDNTFTYTSKASDEDDGFAELYLKMTTL	963
Db	720	EMWRNEQFWVIGVGSAAHLFAVFOGILKVLAGIDNTFTYTSKASDEDDGFAELYLKMTTL	779
OY	964	LIPPTLLILVNLGVYAVGVASTAINSGVSWGKPLGKLFPAFWIYHLIPLFLKGLMGRNR	1023
Db	780	LIPPTLLIILNLGVYAVGVASTAINSGVSWGKPLGKLFPAFWIYHLIPLFLKGLMGRNR	839
OY	1024	TPTIYVWMSVLLASIFSLLMWIRIDPFTSRVGTPOILEGGINC	1065
Db	840	TPTIYVWMAILLASIFSLLMWIRIDPFTTRVGTPOITGGINC	881
RESULT 4			
	US-09-900-237-26		
	Sequence 26 Application US/09900237		
	Patent No. US20020120124A1		
	GENERAL INFORMATION:		
	APPLICANT: Allen, Stephen		
	TITLE OF INVENTION: Plant Cellulose Synthases		
	FILE REFERENCE: B81170 US CIP		
	CURRENT APPLICATION NUMBER: US/09/900,237		
	CURRENT FILING DATE: 2001-07-06		
	PRIOR APPLICATION NUMBER: 60/092,844		
	PRIOR FILING DATE: 1998-07-14		
	PRIOR APPLICATION NUMBER: PCT/US99/15871		
	PRIOR FILING DATE: 1999-07-13		
	PRIOR APPLICATION NUMBER: 09/720383		
	PRIOR FILING DATE: 2000-12-21		
	NUMBER OF SEQ ID NOS: 33		
	SOFTWARE: Microsoft Office 97		
	SEQ ID NO 26		
	LENGTH: 1091		
	TYPE: PRT/		
	ORGANISM: Impatiens balsamita		
	US-09-900-237-26		
	Query Match	70.5%; Score 3999.5; DB 10; Length 1091;	
	Best Local Similarity	69.5%; Pred. NO. 0;	
	Matches	750; Conservative 126; Mismatches 162; Indels 41; Gaps 14;	
OY	10	KPKMKNIYQTOQICSDNGKTVYDGRFACIDICSPVPCPYEYERAKGNSCPQCKTRY	69
Db	31	KPLKRNANOQICQIDGTWGSATGDTFVACNCCGPPVPCPYEYERKKGNDCCPQCKTRY	90
OY	70	KRLKSPAIPIGDKEDGLADEGTVFENYPOREKISERMLGWLITGKEEMGEPOYDKREV	129
Db	91	KRKQKSPVEDEDEEDV-DLLENFNPISGKCKNKK-----VTTARRPQGD-QDDIEL	143
OY	130	S-----HNH-----LPRLTSRQDTSGEFSASPERLVSSTIAG-----GKRLPYSSDVNOS	176
Db	144	SVSSSRHDESOQPVALLTHGHSVSGEI--PTPDNHSIRTSQIGIPVEKSIPIYIDPROPV	201
OY	177	PNRRIVDP-----VGLGNVAMKREYDQMKKAKQENKTPVSTIQASSEGGVDIDASTIOL	230
Db	202	AVRIIVDSKDLNSYGLDANVMKREVEGWMKLQEKNNVMQNSRYPEGGG--DTEGTGSG	259



[illegible]

```

; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 1148
; TYPE: PRT
; ORGANISM: Zea mays
US-09-900-237-4

Query Match      69.1%; Score 3922; DB 10; Length 1148;
Best Local Similarity 67.2%; Pred. No. 0;
Matches 734; Conservative 141; Mismatches 161; Indels 56; Gaps 17

QY      4 EGETAGCKPMKNTVPTQTCICSDNNKGTGVYGDGFVADICDSPRCRCYERERDGNOSCP 63
Db      82 DGEPPKPPDQNRNGVCCQICGDVDGNNPGEPPVACBCAFPCRCRCYEREREGNCP 141
QY      64 QCKTRKRLKSGPALPGDKDEGLADEGVFEFNPYOK--EKISEMLGMHLTRGGEEM 120
Db      142 QCKTRFKRLKGCARYPGDEEDGV--DLENEFNWSKNDHSQYLAESEMLAHMSYGGADL 200
QY      121 -GEOPYDKVSHNHLPRILTSRQ--DTSGEFSASPERLSVSTAGKR--LPYSSVD 173
Db      201 DGPQPFHPHP--NVPLLTNGQMVDDIPDDQHALVP-----SVGSGGGRINHPLP--ADP 252
QY      174 NQSPRRRLVP-----VGLGVANKERDQGMKMKQCKNTGVSPTQASBERGVDDIASI 227
Db      253 NLPVQPRSDPDKDLAAYGVGVANKERMSMKQKER-----MGTQNDGSGDDGDD-- 305
QY      228 DLADEALLNDARQPLSRKSVIPSSRINPVRVMTLRLVILFLHYRTFNPVPAFAL 287
Db      306 ---ADLPPLM--DEAKRQPLSKRITLPSQINPYNKIIIRLVLCFFHHYKVMHNPAPAF 361
QY      288 MLVSYICETIWFALSMILDQPPKMFVYNRETYLDRLALRYDRGEPSOLAVIDIESTVDP 347
Db      362 MLISYICEIWFAMSMILDQPKMFPITERETYLDRLSLRDKEGHPSQLAPVDFESTVDP 421
QY      348 LKEPRLVANTVYLSIADVPRVDKYSCTYFDGGAAMLSTESLAETSEFARKVYRFECKYS 407
Db      422 LKEPRLVANTVYLSIADVPRVDKYSCTYFDGGAAMLSTESLAETSEFARKVYRFECKYS 481
QY      408 IERRADENWEFAKIDYLDKQVOTSPFKDQRAKREVEERKIRNALVYSALKCREGVMV 467
Db      482 LERRADENWFOOKIDYLDKQVAPNFEYKERRAKREVEERKIRNALVYAAQKPRDEGWTY 541
QY      468 QDGTWMPGNNTDHEGMIQVFLGONGGLDAEGNELPRLYVVSREKRRPGEONHKKAGAMA 527
Db      542 QDGTWMPGNNVADHGMIOVFLGOSGHDVEGELPRLYVVSREKRRPYNHKKAGAMA 601
QY      528 LVRSASVLTNGEFLINDCDHATYNNSKALREAMCELMRNLGKQVCTVORPQRFQIDKN 587
Db      602 LVRSASVLTNAYLYLTINDCDHATYNNSKALKEAMCEFMMDLGLKQVCTVORPQRFQIDRH 661
QY      588 DRYARNRNVFEPDINLRGLDGTQGPVYVVGCPFNRLALGYEPRILVKKRKP----- 640
Db      662 DRYARNRNVFEPDINNRGLDGTQGPVYVVGCPFNRLALGYDAP--KTKRPPSRPCNC 718
QY      641 ---LLSKLGGGGRKKRNSKAKKESDK---KSGRHTDSTVYFNLDDIEGEGVAFDE 693
Db      719 PKWICCCCGGRKKRKKTKTKTSKPRFEKIKKLFEKKKNQAPAYALGEIDEAARGA--ENE 776
QY      694 KALLMSQSLERKFGQASAVASTLMENGQVPPSATPENLKEALHIVISCGYEDKSDKMG 753
Db      777 KASIVNOQKLEKKFGQSSVFAVASTLLENGQVPPSATPENLKEALHIVISCGYEDKSDKMG 836
QY      754 EIGWTYGVSTEDILTGFKMHAAGWSTYCMPLRPFKGSAPINLSRLNOVLRYMALGSVE 813
Db      837 DIGWTYGVSTEDILTGFKMHCWGWSYCIPIPKRAAKGSAPINLSRLNOVLRYMALGSJE 896
QY      814 ILFSHCPHWYQNGRLKLEFFAAVNTYITPSIPILMYCTSLAVCFETNPQIPQIS 873
Db      897 IFSNHCPHWYQGGGKLFLEFSTINSIVYWTSPILALYCTLPAILCLDGTGKTFIPELN 956
QY      874 NIASIMPLSLFLSIPATGILLEMRSVGVIDEWMRNEDQVNVIGVSAHLEPAVFGILKVLA 933

```

Db 957 NVASLWFSJLFCIFAFSILEMRWSGVGIDDMRREOFVWIGVSSHLFAVFOGLKVIA 1016  
Qy 934 GIDNFWTSKASDDEGDAELYLKFWTLLIPPTLLIIVNLGVVAIVSAINSQYSW 993  
Db 1017 GVDISFVTSKGGD-DEESELTYFKWTTLLIPPTLLIIVNLGVVAISNAINSGYSW 1075  
Qy 994 GFLGKLEFARVIVHLXPLFKGLMGKRONRPTPIVWVSVLASIFSLMVRIDPFTSRV 1053  
Db 1076 GFLGKLEFARVIVHLXPLFKGLMGKRONRPTPIVWVSVLASIFSLMVRIDPFLAKD 1135  
Qy 1054 TGPDILECGINC 1065  
Db 1136 DGPLLEECGLDC 1147

RESULT 6  
US-09-900-237-10  
; Sequence 10, Application US/09900237  
; Patent No. US20020120124A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen  
; TITLE OF INVENTION: Plant Cellulose Synthases  
; FILE REFERENCE: BBI170 US CIP  
; CURRENT APPLICATION NUMBER: US/09/900, 237  
; CURRENT FILING DATE: 2001-07-06  
; PRIOR APPLICATION NUMBER: 60/092,844  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: PCT/US99/15871  
; PRIOR FILING DATE: 1999-07-13  
; PRIOR APPLICATION NUMBER: 09/720383  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 10  
; LENGTH: 1086  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-900-237-10

Query Match 66.3%; Score 3875.5; DB 10; Length 1086;  
Best Local Similarity 66.7%; Pred. No. 0;  
Matches 728; Conservative 145; Mismatches 161; Indels 57; Gaps 18;

Qy 4 EGERTAKPMKNIIVPOTQICSDNKGTVGDRFAACDICSFPYKPCYEYERKQNSQCP 63  
Db 23 DGDGPPRPQNGQVQCICGDDVGLAPGDPFVACNCAFPVCDCEYERREGTONCP 82  
Qy 64 QCKTRYKRLKSGPAIPGDKDEGLADECTVEFNYP--QEKISERMLGMHLTRGK-GEEM 120  
Db 83 QCKTRYKRLKGCQVATDEEDGY-DLDNDEFNMDGHDSQVASEMLGHNISYGRGDPN 141  
Qy 121 GEPQYDEKSHNLRLPRLTSRQ--DTSGEFSASPERLSVSTJAGKR---LPPSSDVN 174  
Db 142 GAPAQFQ--LNPVNPVLLTNGQVMDIPPEQHALVPSFG-----GGKRIRHPLRY-ADPS 193  
Qy 175 QSPRRRIYDP-----VGLGNVANKERYDGMKMKQEKMTGCVSYQAASERGVDDIDASTD 228  
Db 194 LPVQPRSDPSPKDLAAYGVGSVANKEREMNKQKER-----MHGTGMDGGDDDD--- 245  
Qy 229 ILADEALINDAROPLSRKVSIPSSRIINPYRMVJLRYVILCLFLHYITNPVPNAPALM 288  
Db 246 --ADLPLM-DEARQOLSKRIPLPSSQINPRMIIIRLVVGLGFFHYVMHPVNDAPALM 302  
Qy 289 LVSVICETWALSWILDOFPKWPFPVNRETYLDRLALRYDRGEPQSOLAVIDFVSTVDPL 348  
Db 303 LISVICEIWFMSWILDOFPKWPFPRIERTYLDRLSLRDKGQPSQLAPIDFVSTVDPL 362  
Qy 349 KEPPLVANTVLSLIVADYVYDKVSCYVFDGAAMLSPESLAESEFARKVPPCKKYSI 408  
Db 363 KEPPLVNTVNTVLSLIVADYVYDKVSCYVSDGAAMLFEALSETSEFARKVPPCKKRYNI 422  
Qy 409 EPRAPENYFAAKIDYLDKDVQTSFVKDRRAKREYEERKIRINALYSKALCPREGVWQ 468

Db 423 EPRAPENYFOOKIDYLDKDVANANFVERRRAMREYEERKIRINALYAKAOKVPEEGVTMQ 482  
Qy 469 DGTIPKGNNTGDHPMIOVFLQNGGLAEGNELPRLVYVSEKRPQGHKKAKAMNL 528  
Db 483 DGTIPKGNVNRDHPMIOVFLQNGGLAEGNELPRLVYVSEKRPQGHKKAKAMNL 542  
Qy 529 VRVSAVLNPGFILLDDCDHYINNSKALREAMCFMDNGLKQVGYVOPORFSDIDKND 588  
Db 543 VRVSAVLNAPILLDDCDHYINNSKALREAMCFMDNGLKQVGYVOPORFSDIDKND 602  
Qy 589 RYANRNTVFPDINLGLDGIQGVYVTCGVFNRTALYGEPPYIKKHKPSLSKLCG- 647  
Db 603 RYANRNTVFPDINLGLDGIQGVYVTCGVFNRTALYGEPPYIKKHKPSLSKLCG- 657  
Qy 648 -----GSRKKNKAKKESDKR--SGRHTDSVYPPVFNLDIDIEGVEGAFDDEK 694  
Db 658 WPKWCFCCCGENRRQKTKRKTREKRLFEKKEENOSPVALGEIDEAALPGA--ENEK 715  
Qy 695 ALLMSOMSLERKFGQSAFVASTLMENGVPPSATPENLKEAHIVISGYEDKSDMGME 754  
Db 716 AGIVNQQLEKKFGQSSVFVSTLLENGGTLKASAPASLKEAHIVISGYEDKSDMGME 775  
Qy 755 IGWISYTEDILITGFKMHARGRSIYCPKLPARFGSAPINLSDRLNOVLRMALGSVEI 814  
Db 776 IGWISYTEDILITGFKMHARGRSIYCPKLPARFGSAPINLSDRLNOVLRMALGSVEI 835  
Qy 815 LFSRHCPITWYNGNKLKLERFAYVNTTITPTISPLMYCTLAVCLFTNOFTIPIQISN 874  
Db 836 FFSNHCPLMYGSGGLKFLERFSYNSIYVPTWSIPLAYCTLPACILCTGKFTPELNN 895  
Qy 875 IASIFLPLSIFATGILKEMWSGVGIDEMMRNQPWVIGVSAHLFAVFOGLKVIA 934  
Db 896 VASLWFSJLFCIFAFSILEMRWSGVGIDEMMRNQPWVIGVSAHLFAVFOGLKVIA 955  
Qy 935 IDNFWTSKASDDEGDAELYLKFWTLLIPPTLLIIVNLGVVAIVSAINSQYSW 994  
Db 956 VDISFTVSKGGD-DEESELTYFKWTTLLIPPTLLIIVNLGVVAIVSAINSQYSW 1014  
Qy 995 PFLGKLEFARVIVHLXPLFKGLMGKRONRPTPIVWVSVLASIFSLMVRIDPFTSRV 1054  
Db 1015 PFLGKLEFARVIVHLXPLFKGLMGKRONRPTPIVWVSVLASIFSLMVRIDPFLAKD 1074  
Qy 1055 GPDILECGINC 1065  
Db 1075 DGPLLEECGLDC 1085

RESULT 7  
US-09-900-237-8  
; Sequence 8, Application US/09900237  
; Patent No. US20020120124A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen  
; TITLE OF INVENTION: Plant Cellulose Synthases  
; FILE REFERENCE: BBI170 US CIP  
; CURRENT APPLICATION NUMBER: US/09/900, 237  
; CURRENT FILING DATE: 2001-07-06  
; PRIOR APPLICATION NUMBER: 60/092,844  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: PCT/US99/15871  
; PRIOR FILING DATE: 1999-07-13  
; PRIOR APPLICATION NUMBER: 09/720383  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 8  
; LENGTH: 1165  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-900-237-8

Query Match 67.7%; Score 3843.5; DB 10; Length 1165;



[illegible]

```

RESULT 10
US-09-838-539-6
: Sequence 6, Application US/09838539
: Patent No. US20020129401A1
: GENERAL INFORMATION:
: APPLICANT: Stalker, D. et al.
: TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
: TITLE OF INVENTION: Sequences
: FILE REFERENCE: 15621/03/US
: CURRENT APPLICATION NUMBER: US/09/838,539
: CURRENT FILING DATE: 2001-04-18
: PRIOR APPLICATION NUMBER: 60/029,987
: PRIOR FILING DATE: 1996-10-29
: PRIOR APPLICATION NUMBER: 08/960,048
: PRIOR FILING DATE: 1997-10-29
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 974
: TYPE: PRT
: ORGANISM: Gossypium hirsutum
US-09-838-539-6

```

```

0y      12  MNIVPOTQICSDNVGKTVGDGRFVADICISFPYCRRCITYEKERKDNQSGPCKCTTKRR  71
      1  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db      2  MESSVP-VCHRCGGHVLNAGGEPRVADHCENLFPKICSCFEPYDLKEGKRAQLRC-----  54
      1  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Query Match 61.2%, Score 3473.5, DB 10, Length 974,
Best Local Similarity 62.1%, Pred. No. 4,3e-290,
Matches 660, Conservative 129, Mismatches 176, Indels 97, Gaps 15,

```

72 LKSPALPDKDEGLADGCTVEFNPPOKEKISERNLGHHLTRGKEGEBPOYDKEVSH 131  
55 -GSP-----YDENLDD-----VERATGDQSTMAHL--NKSDVG-----HIA 90  
132 NMLPRLSHQDTSGEESASPERLSVSTIAGKRLLPYSSDVQSPNRRIYDVGIGNA 191  
91 RHISVSSTJLDSKAEEN-----GNSI 111  
192 WKERVDMKMOKEKNTGVSTQASERGVGVIDASTDILADEALLND---EAKOPLSRK 247  
112 MKRWESMEKKKKKKPATTKVERAE-----IPPEQOMDKPADAPASQCLSTI 161  
248 VSPSSRINPRKMYMLRLVILCLPLHYITNPVNAFLMLVSVCEIMFALSWLDDP 307  
162 IPIPSRLAPYRVLIIIMRLIILGFPHYTNVPDASAFGLMTLSVICIEMFAESWLDDP 221  
308 PKMFVNRRTYDLRLALRYDREGESQLAADVLFVSTVPLKEPPLVANTVLSIADY 367  
222 PKMYPNRRTYIDRLSAREGERGEPDLAAVDFVSTVPLKEPPLITANTVLSIALDY 281  
368 PVDKVSICYVDDGAAMLSESLAETSEFARKWVPFCCKYSIEBRAPEMYFAAKIDYLDK 427  
282 PVDKVSICYSDGGAAMLTFESLVEETADFAKWPVFCCKFSIEBRADEFYSOKIDYLDK 341  
428 VQTSFYKDBRARKREEEKIRINALYSALKCPREBGMWODSTPMRGNTGDHPGMIOY 487  
342 VQTSFYKDBRARKREEEKIRINALYSALKCPREBGMWODSTPMRGNTGDHPGMIOY 401  
488 FFGONGGLDAEGNELPRLVYVSEKRGPGOHKKKAGAMALVSVAVLTPNGPILDDC 547  
402 FFGYSGARDIEGNEPRLVYVSEKRGPGOHKKKAGAMALVSVAVLTPNGPILDDC 461  
548 HTNNSKALREAMCFILMDPILKQVCYVOPORFDGIDKNDRYANRNTVFEDINLGLDG 607  
462 HTNNSKALREAMCFILMDPILKQVCYVOPORFDGIDKNDRYANRNTVFEDINLGLDG 521  
608 IGGPVYVGTGVENRATLGYEPPILKVRKKPSILSKLGGSKKKSKAKSKSDKKSR 667  
522 IGGPVYVGTGVENRATLGYEPPILKVRKKPSILSKLGGSKKKSKAKSKSDKKSR 579  
668 HTDSTVPVNLDDIEGVEGAGFDD-EKALLMSOMLEKRFQSAVFAVSTLMENGVP 726  
580 REELDAIFNLREID-----NYDEYERMLISQTSFEKTFGLSSVFISTLMENGVAE 633  
727 SATPENLKEAIVHISCGYEDKSDMGMEIGWIGSVTEIDLTFGKMHARGRSIYCMPLR 786  
634 SANPSLLKEAIVHISCGYEDKSDMGMEIGWIGSVTEIDLTFGKMHARGRSIYCMPLR 693  
787 PAFKGSAPINLSDRLNOVLRLMALGSVEILFSRHCPITWGY-NGRKFLERFAVNTIYP 845  
694 PAFKGSAPINLSDRLNOVLRLMALGSVEILFSRHCPITWGY-NGRKFLERFAVNTIYP 753  
846 IASIPILMYCTLLAVCLFTNOFTIPIQISNIASIFLSTLFIATGILMRMSGVIGDM 905  
754 FLSPLIACSLPACILTLGKFLIPTLSLAVSLFGLFSLITVAVLELRMSGVIEDL 813  
906 WNEBQFVIGVSAHLFAVFOGILKVLAGIDINFTVTSKASDEGDFAELYLEFKWTTLLI 965  
814 WNEBQFVIGVSAHLFAVFOGILKVLAGIDINFTVTSKASDEGDFAELYLEFKWTTLLI 872  
966 PPTLLIIVLVGVVGVSAINSQVOSMPLFGKLFEPFVAVIHLVLPFLKGLMGQRNRP 1025  
873 PPTLLIIVLVGVVGVSAINSQVOSMPLFGKLFEPFVAVIHLVLPFLKGLMGQRNRP 932  
1026 TIVVWVSVLASIFSLWVRIDPFTSRVYGPDLLE--CGINC 1065  
933 TIVVWVSVLASIFSLWVRIDPFTSRVYGPDLLE--CGINC 974

RESULT 11  
US-09-900-237-18  
; Sequence 18, Application US/09900237  
; Patent No. US20020120124A1

GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen  
; TITLE OF INVENTION: Plant Cellulose Synthases  
; FILE REFERENCE: B01170 US CIP  
; CURRENT APPLICATION NUMBER: US/09/900,237  
; PRIOR FILING DATE: 2001-07-06  
; PRIOR APPLICATION NUMBER: 60/092,844  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: PCT/US99/15871  
; PRIOR FILING DATE: 1999-07-13  
; PRIOR APPLICATION NUMBER: 09/720383  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO: 18  
; LENGTH: 793  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-09-900-237-18

Query Match 56.6%; Score 3216; DB 10; Length 793;  
Best Local Similarity 72.3%; Pred. No. 4,8e-268;  
Matches 573; Conservative 118; Mismatches 90; Indels 12; Gaps 5;

279 NPYNNAFALMLVSVCEIMFALSWLDDPFPKFPVNRRTYDLRLALRYDREGESQLA 338  
4 HPYNDAYGLMLTSVCEIMFAVSWIMDDPKWYPIORETLYDLRLSLRYEKKPSLSV 63  
339 DIFVSTVDPDLKEPPLTANTVLSIADYVPOVKSVCYVDDGAAMLSESLAETSEFARK 398  
64 DIFVSTVDPDLKEPPLTANTVLSIADYVPOVKSVCYVDDGAAMLSESLAETSEFARK 123  
399 WVPFCCKYSIEBRAPEMYFAAKIDYLDKQVTSFYKDBRARKREEEKIRINALYSKAL 458  
124 WVPFCCKYSIEBRAPEMYFAAKIDYLDKQVTSFYKDBRARKREEEKIRINALYSKAL 183  
459 KCPREBGMWODSTPMRGNTGDHPGMIOYFLQONGGLDAEGNELPRLVYVSEKRGPGOH 518  
184 KVPEDWMTQDGTPMWGNVNRDHPGMIOYFLQONGGLDAEGNELPRLVYVSEKRGPGOH 243  
519 HKKAGAMALVSVAVLTPNGPILNDCHYINNSKALREAMCFILMDPILKQVCYVOP 578  
244 HKKAGAMALVSVAVLTPNGPILNDCHYINNSKALREAMCFILMDPILKQVCYVOP 303  
579 QRFDGIDKNDRYANRNTVFEDINLGLDGIOGPVYVGTGVENRATLGYEPPILKVRKK 638  
304 QRFDGIDKNDRYANRNTVFEDINLGLDGIOGPVYVGTGVENRATLGYEPPILKVRKK 363  
639 -----PILLSKLCGSSKRSKAKSKESKRSKGRITDSTVAVENLDDIEGVEGAGFDD 692  
364 KTCNCMPKWCCLCCGSRKKRNKNSKE--KKRKYHSEASKOIHALENTAAGNEGT--NN 419  
693 EKALLMSOMLEKRFQSAVFAVSTLMENGVPASATPENLKEAIVHISCGYEDKSDMG 752  
420 EKTSNLTQKLEKRGQSPVFAVSTLLDDGVPFHGVSPLSLKEAIOVISCYEDKTEWG 479  
753 MEIGWIGSVTEIDLTFGKMHARGRSIYCMPLFAKGSAPINLSDRLNOVLRLMALGSV 812  
480 KEVGMWIGSVTEIDLTFGKMHARGRSIYCMPLFAKGSAPINLSDRLNOVLRLMALGSV 539  
813 ELFSRHCPITWGY-NGRKFLERFAVNTIYPISIPILMYCTLLAVCLFTNOFTIPIQI 872  
540 ELFSRHCPITWGY-NGRKFLERFAVNTIYPISIPILMYCTLLAVCLFTNOFTIPIQI 599  
873 SNIASIFLSTLFIATGILMRMSGVIGDMWNEBQFVIGVSAHLFAVFOGILKVL 932  
600 SNIASIFLSTLFIATGILMRMSGVIGDMWNEBQFVIGVSAHLFAVFOGILKVL 659  
933 AGIDINFTVTSKASDEGDFAELYLEFKWTTLLIPTLLIIVLVGVVGVSAINSQVOS 992  
660 AGVNTVFTVSKAAD-DGEFSELYTFKWTSLIIPMTLLIIVLVGVVGVSAINSQVOS 718  
993 WGPLFGKLFEPFVAVIHLVLPFLKGLMGQRNRPITIVVWVSVLASIFSLWVRIDPFTSR 1052

Db 719 WGPLGRLEFALWVILHLPFLKGLGKODRMPITLWVSIILSLITLWVRINPEVSR 778  
QY 1053 VTGPDLIEGICINC 1065  
Db 779 -DGPVLEICGLNC 790

RESULT 12  
US-09-900-237-24  
; Sequence 24, Application US/09900237  
; Patent No. US20020120124A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen  
; TITLE OF INVENTION: Plant Cellulose Synthases  
; FILE REFERENCE: Bb1170 US CIP  
; CURRENT APPLICATION NUMBER: US/09/900,237  
; CURRENT FILING DATE: 2001-07-06  
; PRIOR APPLICATION NUMBER: 60/092,844  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: PCT/US99/15871  
; PRIOR FILING DATE: 1999-07-13  
; PRIOR APPLICATION NUMBER: 09/720383  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 24  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Plectramnia pentandra  
US-09-900-237-24

Query Match  
Best Local Similarity 52.9%; Score 3003 5; DB 10; Length 740;  
Matches 547; Conservative 94; Mismatches 90; Indels 13; Gaps 5;

QY 338 RGEPSOLAANDIFVSTVDPLKEPPLVANTVLSILAVDYPVCKVSCYFDDGAAMLSPFE 387  
Db 3 KEGKPSLGLADIDFVSTVDPMKEPPLITANTVLSILAVDYPVCKVSCYFDDGAAMLSPFE 62

QY 368 SLAESEFARKKVPCKKXISTEPRAPEWTFKAKIDYLDKDVOTSFVKDRAMKREYEERK 447  
Db 63 ALSESEFARKKVPCKKXISTEPRAPEWTFKAKIDYLDKDVOTSFVKDRAMKREYEERK 122

QY 448 IRLNLYSKALCKPREGVWMDGTPMPCNNTGDHPGMIQVFLGONGGLDAEGNELPRLVY 507  
Db 123 VRINKLVAMAQKVPEDGTMODGTPMPCNNTGDHPGMIQVFLGONGGLDAEGNELPRLVY 182

QY 508 VSREKRRPGFOHKKKAGANALVRSVAVLTNGPFIINLDCDHYINNSKALREAMCFMLMDPN 567  
Db 183 VSREKRRPGFOHKKKAGANALVRSVAVLTNGPFIINLDCDHYINNSKALREAMCFMLMDPN 242

QY 568 LGKQVCYVQFQOREGIDKNDRYANRNTVFEDINLRGIDGLOGPYVGTGCVFNRTALYG 627  
Db 243 SGKKLCYVQFQOREGIDKNDRYANRNTVFEDINLRGIDGLOGPYVGTGCVFNRTALYG 302

QY 628 YEPFLKVVHKK-----PSLSKLCGSGRKKNSKAKKESDKKSGRHTSDTVPFVFNLDI 681  
Db 303 YDAPVTKKSPGKACMCWPKMLCCCGGSRKKNSKAKKESDKKSGRHTSDTVPFVFNLDI 359

QY 662 BEGVGAGFDDKALLMSOMSLKRRGOSAVFVASTLMEGNGVPPSAPPENILKLEAIIYI 741  
Db 360 BEGM--GGINSEKSETTPPLKLEKKEGSPVFAVASTLLEDGVPDQATPAALLKEIYI 417

QY 742 SCGYEDKSDMGMEICMTIGSVYEDILITGFKMHARMSRYICMPKLPARFGSAPINISDRL 801  
Db 418 SCGYEDKSDMGMEICMTIGSVYEDILITGFKMHARMSRYICMPKLPARFGSAPINISDRL 477

QY 802 NOVLRMALGSVILSRHCPIMYGNGRLKFLERFAVYNTTITPITSIPLLWYCTLLAC 861  
Db 478 NOVLRMALGSVILSRHCPIMYGNGRLKFLERFAVYNTTITPITSIPLLWYCTLLAC 537

QY 862 LETNOFLIPQISINIASIMFLSLFISFATGILEMRSGVIGIDEMWRNEQFVWIGVSAHL 921

Db 538 LTKGRFIVEISINVASIILEMFLIFIAITSILEMQGSGVIGIDEMWRNEQFVWIGVSAHL 597  
QY 922 FAVFOGILKVLGADIDNFTVAKASDEDDDEALYLFKKTTLIPPTILLIYLVAVG 981  
Db 598 FALFOGLKVLGAVNTNFTVAKADE--GDFSELVLFKKTTLIPPTILLIYLVAVG 656

QY 982 VSAVINSYGOSWGPLEGLFFAFWVIVHLXYPFLKGLMGQRNRPRTIVVWVSLASISFL 1041  
Db 657 VSDALINGYDSNGPLFCGLFFAFWVIVHLXYPFLKGLMGQRNRPRTIVVWVSLASISFL 716

QY 1042 LWRIDPFTSRVTPDLIEGICINC 1065  
Db 717 LWVRINPEVSR--DGPVLEICGLNC 739

RESULT 13  
US-09-900-237-31  
; Sequence 31, Application US/09900237  
; Patent No. US20020120124A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen  
; TITLE OF INVENTION: Plant Cellulose Synthases  
; FILE REFERENCE: Bb1170 US CIP  
; CURRENT APPLICATION NUMBER: US/09/900,237  
; CURRENT FILING DATE: 2001-07-06  
; PRIOR APPLICATION NUMBER: 60/092,844  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: PCT/US99/15871  
; PRIOR FILING DATE: 1999-07-13  
; PRIOR APPLICATION NUMBER: 09/720383  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 31  
; LENGTH: 685  
; TYPE: PRT  
; ORGANISM: Gossypium hirsutum  
US-09-900-237-31

Query Match  
Best Local Similarity 49.5%; Score 2808; DB 10; Length 685;  
Matches 512; Conservative 77; Mismatches 77; Indels 22; Gaps 6;

QY 397 RKMVPCKKXISTEPRAPEWTFKAKIDYLDKDVOTSFVKDRAMKREYEERKIRNALVSK 456  
Db 1 RKMVPCKKXISTEPRAPEWTFKAKIDYLDKDVOTSFVKDRAMKREYEERKIRNALVSK 60

QY 457 ALCKPREGVWMDGTPMPCNNTGDHPGMIQVFLGONGGLDAEGNELPRLVYVSREKRRPG 516  
Db 61 AOKPREGVWMDGTPMPCNNTGDHPGMIQVFLGONGGLDAEGNELPRLVYVSREKRRPG 120

QY 517 OHKKKAGANALVRSVAVLTNGPFIINLDCDHYINNSKALREAMCFMLMDPNLGKQVCYVQ 576  
Db 121 OHKKKAGANALVRSVAVLTNGPFIINLDCDHYINNSKALREAMCFMLMDPNLGKQVCYVQ 180

QY 577 FQOREGIDKNDRYANRNTVFEDINLRGIDGLOGPYVGTGCVFNRTALYGPERPKVNH 636  
Db 181 FQOREGIDKNDRYANRNTVFEDINLRGIDGLOGPYVGTGCVFNRTALYGPERPKVNH 240

QY 637 KR-----PSLSKLCGSGRKKNSKAKKESDKKSGRHTSDTVPFVFNLDI 678  
Db 241 KR-----PSLSKLCGSGRKKNSKAKKESDKKSGRHTSDTVPFVFNLDI 300

QY 679 DDIEGVGAGFDD--EKALLMSOMSLKRRGOSAVFVASTLMEGNGVPPSAPPENILKLEA 737  
Db 301 DDIEGVGAGFDD--EKALLMSOMSLKRRGOSAVFVASTLMEGNGVPPSAPPENILKLEA 358

QY 738 IIVISGCVYEDKSDMGMEICMTIGSVYEDILITGFKMHARMSRYICMPKLPARFGSAPINL 797  
Db 359 IIVISGCVYEDKSDMGMEICMTIGSVYEDILITGFKMHARMSRYICMPKLPARFGSAPINL 418

QY 798 SDRLNOVLKMALGSVILSRHCPIMYGNGRLKFLERFAVYNTTITPITSIPLLWYCTLL 857

Db 419 SRELHOVLKRALGSAVEIFLSRHCPIMYGYGKGLKMLERLAYINTIYPPFISILLACYT 478  
QY 858 LAVCLETNOFIIPQISNIASIMFSLFSIFATGILEMRSGVIGDEMNEQFVIGV 917  
Db 479 PAVCLITGFIIPITLSNLTSMVFLALFSLIATGVLRLMRSVSIODMMRNEQFVIGV 538  
QY 918 SAHLFAVFOGILKVLAGIDTNTFTVTSKASDEGDAELYLKFWTLLIPTLLIYNLVG 977  
Db 539 SAHLFAVFOGILKVLAGVDTNFTVAKAD-DTEFELYLKFWTLLIPTLLIYNLVG 597  
QY 978 VVAAGSYAINSGYOSMGPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIIVVWSVLLAS 1037  
Db 598 VVAAGSDAINNGSGMPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIIVVWSVLLAS 657  
QY 1038 IFSLLMVRIDPFTSRVTPGDIIECGI 1065  
Db 658 IFSLLMVRIDPFTSRVTPGDIIECGI 685

## RESULT 14

US-09-838-539-7  
; Sequence 7, Application US/09838539  
; Patent No. US20020129401A1  
; GENERAL INFORMATION:  
; APPLICANT: Stalker, D. et al.  
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter  
; FILE REFERENCE: 15621/03/US  
; CURRENT APPLICATION NUMBER: US/09/838,539  
; CURRENT FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/029,987  
; PRIOR FILING DATE: 1996-10-29  
; PRIOR APPLICATION NUMBER: 08/960,048  
; PRIOR FILING DATE: 1997-10-29  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 685  
; TYPE: PR  
; ORGANISM: Gossypium hirsutum  
US-09-838-539-7

Query Match 49.4%; Score 2803; DB 10; Length 685;  
Best Local Similarity 74.5%; Pred. No. 1.5e-232;  
Matches 512; Conservative 77; Mismatches 76; Indels 22; Gaps 6;

QY 396 ARKWPECKKYSIEPAPFMYPAKTDYLDKDYQTSFVADRDRAMKREYEFKIRIATLYS 455  
Db 1 ARKWPECKKHNVEPRAPFEYFENEKIDYLDKDYHPSFVEERRAMKREYEFKIRIATLYS 60  
QY 456 KALKCEEGWVWODGTPMPGNNTPGDPHGMIOVFLGONGGLDAEGNELPRLVYVSREKRP 515  
Db 61 KAKKREBEWVWODGTPMPGNNTRDHPGMIOVFLGASAGALDVGKELPRLVYVSREKRP 120  
QY 516 FOHHRKAGAMNALIVRSVAVLTNGPFLTLNLDCHYINNSKALREAMCFELMDPNLGKQCVY 575  
Db 121 YOHHRKAGAMNALIVRSVAVLTNAPFLTLNLDCHYINNSKALREAMCFELMDPNLGKQCVY 180  
QY 576 OFPQRFDDIDKDRANRNTVFEDINLRGLDGIQGVYVYVGGCVNFRALYGEPRITVY 635  
Db 181 OFPQRFDDIDKDRANRNTVFEDINLRGLDGIQGVYVYVGGCVNFRALYGEPRITVY 240  
QY 636 HKK-----PSLISKLKSGSKRSKRSKAKE-----SDKRSGRH--TDSVTPVN 677  
Db 241 RKMTCDCKPMSMCCCCGSGSKRSKRSKKEKGLLGLLIGKKKMMKKNVYKGSAPVD 300  
QY 678 LDDIEGVGAGFD- EKALLMSQMSLEKRFQSAVAVASTLMENGVPSPATPENILKE 736  
Db 301 LDDIEGVGAGFD- EKALLMSQMSLEKRFQSAVAVASTLMENGVPSPATPENILKE 358  
QY 737 AIHVISCGYEDKSDMGMELGWTIGSTYEDTILTFKMHARGMSIYMPKLPARFGSAPIN 796  
Db 796 AIHVISCGYEDKSDMGMELGWTIGSTYEDTILTFKMHARGMSIYMPKLPARFGSAPIN 796

Db 359 AIHVISCGYEETKEMKEIGWTIGSTYEDTILTFKMHARGMSIYMPKLPARFGSAPIN 418  
QY 797 LSRDLNOVLRMALGSAVEIFLSRHCPIMYGNGLKLEBFAVNTTIYPTISPLMAYC 856  
Db 419 LSRDLNOVLRMALGSAVEIFLSRHCPIMYGNGLKLEBFAVNTTIYPTISPLMAYC 478  
QY 857 LAVCLETNOFIIPQISNIASIMFSLFSIFATGILEMRSGVIGDEMNEQFVIGV 916  
Db 479 IPAVCLITGFIIPITLSNLTSMVFLALFSLIATGVLRLMRSVSIODMMRNEQFVIGV 538  
QY 917 VSAHLFAVFOGILKVLAGIDTNTFTVTSKASDEGDAELYLKFWTLLIPTLLIYNLV 976  
Db 539 VSAHLFAVFOGILKVLAGVDTNFTVAKAD-DTEFELYLKFWTLLIPTLLIYNLV 597  
QY 977 VVAAGSYAINSGYOSMGPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIIVVWSVLLA 1036  
Db 598 VVAAGSDAINNGSGMPLFGKLFPAFWIYHLIYPLKGLMGRONRPTIIVVWSVLLA 657  
QY 1037 IFSLLMVRIDPFTSRVTPGDIIECGI 1063  
Db 658 IFSLLMVRIDPFTSRVTPGDIIECGI 684

## RESULT 15

US-09-900-237-16  
; Sequence 16, Application US/09900237  
; Patent No. US20020120124A1  
; GENERAL INFORMATION:  
; APPLICANT: Allen, Stephen  
; TITLE OF INVENTION: Plant Cellulose Synthases  
; FILE REFERENCE: B81170 US CIP  
; CURRENT APPLICATION NUMBER: US/09/900,237  
; CURRENT FILING DATE: 2001-07-06  
; PRIOR APPLICATION NUMBER: 60/092,844  
; PRIOR FILING DATE: 1998-07-14  
; PRIOR APPLICATION NUMBER: PCT/US99/15871  
; PRIOR FILING DATE: 1999-07-13  
; PRIOR APPLICATION NUMBER: 09/720383  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 610  
; TYPE: PR  
; ORGANISM: Glycine max  
US-09-900-237-16

Query Match 46.6%; Score 2647.5; DB 10; Length 610;  
Best Local Similarity 78.8%; Pred. No. 3.1e-219;  
Matches 482; Conservative 55; Mismatches 72; Indels 3; Gaps 2;

QY 455 SKALCEEGWVWODGTPMPGNNTPGDPHGMIOVFLGONGGLDAEGNELPRLVYVSREKRP 514  
Db 1 AKAKPPEBEWVWODGTPMPGNNTRDHPGMIOVFLGSHSGDLTDGELPRLVYVSREKRP 60  
QY 515 GFOHHRKAGAMNALIVRSVAVLTNGPFLTLNLDCHYINNSKALREAMCFELMDPNLGKQCVY 574  
Db 61 GFOHHRKAGAMNALIVRSVAVLTNGAVLTNLDCHYINNSKALREAMCFELMDPNLGKQCVY 120  
QY 575 VOFQRFDDIDKDRANRNTVFEDINLRGLDGIQGVYVYVGGCVNFRALYGEPRITVY 634  
Db 121 VOFQRFDDIDKDRANRNTVFEDINLRGLDGIQGVYVYVGGCVNFRALYGEPRITVY 180  
QY 635 KHKPFLSLKLGSGSKRSKRSKKA-KRSDKRSGRHTDSVTPVFNLDIDIEGVGAGFDE 693  
Db 181 EDLEPITVYSCGSGSKRSKRSKKA-KRSDKRSGRHTDSVTPVFNLDIDIEGVGAGFDE 736  
QY 694 KALLMSQMSLEKRFQSAVAVASTLMENGVPSPATPENILKEAIHVISCGYEDKSDMGM 753  
Db 239 KALLMSQMSLEKRFQSAVAVASTLMENGVPSPATPENILKEAIHVISCGYEDKSDMGM 796  
QY 754 EIGWTIGSTYEDTILTFKMHARGMSIYMPKLPARFGSAPINLQVLRMALGSAVE 813  
Db 813 EIGWTIGSTYEDTILTFKMHARGMSIYMPKLPARFGSAPINLQVLRMALGSAVE 813





THIS PAGE BLANK (USPTO)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: June 16, 2003, 11:05:53 ; Search time 18 Seconds  
(without alignments)  
1740.856 Million cell updates/sec

Title: AAC39336  
Perfect score: 5677  
Sequence: 1 MESEGETAGKPMKNIVPQC.....IDPFTSRVTPGDIIECGINC 1065

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4133.5	72.8	881	US-08-960-048-8	Sequence 8, Appl
2	3473.5	61.2	974	US-08-960-048-6	Sequence 6, Appl
3	2803	49.4	685	US-08-960-048-7	Sequence 7, Appl
4	278	4.9	693	US-08-960-048-11	Sequence 11, Appl
5	245.5	4.3	861	US-08-960-048-12	Sequence 12, Appl
6	236.5	4.2	756	US-08-960-048-10	Sequence 10, Appl
7	228.5	4.0	756	US-07-689-008-2	Sequence 2, Appl
8	228.5	4.0	3031	US-07-689-008-2	Sequence 3, Appl
9	196	3.5	723	US-08-960-048-9	Sequence 9, Appl
10	196	3.5	723	US-08-960-048-9	Sequence 9, Appl
11	124	2.2	357	US-08-119-773-4	Sequence 4, Appl
12	116	2.0	357	US-08-119-773-6	Sequence 4, Appl
13	115	2.0	346	US-08-119-773-5	Sequence 6, Appl
14	115	2.0	357	US-08-119-773-2	Sequence 6, Appl
15	110	1.9	1587	US-09-000-094-46	Sequence 46, Appl
16	107	1.9	686	US-08-768-301-4	Sequence 4, Appl
17	106.5	1.9	1437	US-09-061-400-2	Sequence 2, Appl
18	106.5	1.9	1453	US-09-001-273-2	Sequence 2, Appl
19	106.5	1.9	1453	US-08-843-459A-2	Sequence 2, Appl
20	104	1.8	1167	US-08-485-568A-6	Sequence 6, Appl
21	104	1.8	1167	US-08-590-554A-6	Sequence 6, Appl
22	104	1.8	1167	US-09-184-223-6	Sequence 6, Appl
23	104	1.8	3033	US-07-925-695-5	Sequence 5, Appl
24	103	1.8	1168	US-08-620-717A-9	Sequence 9, Appl
25	102.5	1.8	1871	US-08-694-869-1	Sequence 1, Appl
26	102.5	1.8	1871	US-09-349-546-1	Sequence 1, Appl
27	101.5	1.8	765	US-08-425-061-19	Sequence 19, Appl

28	101.5	1.8	765	US-08-825-886-19	Sequence 19, Appl
29	101.5	1.8	900	US-08-425-061-20	Sequence 20, Appl
30	101.5	1.8	900	US-08-825-886-20	Sequence 20, Appl
31	101.5	1.8	914	US-08-425-061-21	Sequence 21, Appl
32	101.5	1.8	914	US-08-825-886-21	Sequence 21, Appl
33	101.5	1.8	1202	US-08-425-061-22	Sequence 22, Appl
34	101.5	1.8	1202	US-08-825-886-22	Sequence 22, Appl
35	101.5	1.8	1363	US-08-425-061-23	Sequence 23, Appl
36	101.5	1.8	1363	US-08-825-886-23	Sequence 23, Appl
37	101.5	1.8	1852	US-08-425-061-24	Sequence 24, Appl
38	101.5	1.8	1852	US-08-825-886-24	Sequence 24, Appl
39	101.5	1.8	1863	US-08-598-591-2	Sequence 2, Appl
40	101.5	1.8	1863	US-08-480-784-2	Sequence 2, Appl
41	101.5	1.8	1863	US-08-480-784-2	Sequence 2, Appl
42	101.5	1.8	1863	US-08-487-003-2	Sequence 2, Appl
43	101.5	1.8	1863	US-08-483-554B-2	Sequence 2, Appl
44	101.5	1.8	1863	US-08-798-691-2	Sequence 2, Appl
45	101.5	1.8	1863	US-08-798-691-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1					
US-08-960-048-8					
Sequence 8, Application US/08960048C					
Patent No. 6271443					
GENERAL INFORMATION:					
APPLICANT: Stalker, D. et al.					
TITLE OF INVENTION: Plant Cellulose Synthase and Promoter					
FILE REFERENCE: 15621/01/US					
CURRENT APPLICATION NUMBER: US/08/960,048C					
CURRENT FILING DATE: 1997-10-29					
PRIOR APPLICATION NUMBER: 60/029,987					
NUMBER OF SEQ ID NOS: 12					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 8					
LENGTH: 881					
TYPE: PRT					
ORGANISM: (Oryzae sativa)					
US-08-960-048-8					
Query Match					
Best Local Similarity 72.8%; Score 4133.5; DB 4; Length 881;					
Matches 766; Conservative 55; Mismatches 56; Indels 5; Gaps 3;					
QY	188	GNVAMKERYDGMKMKOEKNTGPV---SQQAASERGV--DIDASTDILADEALNDEAROP	243		
DB	1	GNVAMKERYDGMKMKOEKNTGPV---SQQAASERGV--DIDASTDILADEALNDEAROP	60		
QY	244	LSRKVSIPSSRIINRYIMLVLCFLHYRITPNVPAFALVSVICELPALSNI	303		
DB	61	LSRKVPDSSRIINRYIMLVLCFLHYRITPNVPAFALVSVICELPALSNI	120		
QY	304	IDQFKKPPVARETLDRLALRYDREGESQAAVDIRYSTVDPLKEPLVYANVLSL	363		
DB	121	IDQFKKPPVARETLDRLALRYDREGESQAAVDIRYSTVDPLKEPLVYANVLSL	180		
QY	364	AVDYVVDVYSCVDFDGAAMLSFESLAETSEFARKMVPFCKYSTEPRAPERYFAKIDY	423		
DB	181	AVDYVVDVYSCVDFDGAAMLSFESLAETSEFARKMVPFCKYSTEPRAPERYFAKIDY	240		
QY	424	LKDKVOTSPVXDRAMKREYEEFKIRINALVSKALCDEEGVMODGTPWPNNTGDHPG	483		
DB	241	LKDKVOTSPVXDRAMKREYEEFKIRINALVSKALCDEEGVMODGTPWPNNTGDHPG	300		
QY	484	MIQVTLGONGCIGDAGCNELPRLYVYSREKRPFOHKKAGAMNALVRSAYLTNGPFLN	543		
DB	301	MIQVTLGONGCIGDAGCNELPRLYVYSREKRPFOHKKAGAMNALVRSAYLTNGPFLN	360		
QY	544	LDGDHYINNSKALREAMCEFLMDPNNGKQVYQFQORDGIDKNDRYANRNTVFVDINLR	603		

```

Db      361 LQCDHINNSKALREMAFCLMDPNLGRSVYVOPQRFQDIGNDRYANNNTVFEDINR 420
Oy      604 GIDGIGPYVYGVGCYFNRTALYGPPIKVKHKKRSLSLKCGSRKNSKKKSDKK 663
Db      421 GIDGIGPYVYGVGCYFNRTALYGPPIKVKHKKRSLSLKCGSRKNSKKKSDKK 479
Oy      664 KSGRHTDSTVPVFNLDIEGVEGAGFDEKALIMSOMSLKRRFGOSAVFVASTLMENG 723
Db      480 KSNHVDASAVPYNEDIEGVEGAGFDEKALIMSOMSLKRRFGOSAVFVASTLMENG 539
Oy      724 VPPSATPEMLKFAHIVISCGYEDKSDMGMEIGWITGVTEDILTFKMHARGWSITYCM 783
Db      540 VPOSAPESILKEAHIHIVISCGYEDKSDMGMEIGWITGVTEDILTFKMHARGWSITYCM 599
Oy      784 PLKPAKGSAPINLSRLNOVLRMALGSVEILSRHCPIMWYGNGLKFLERAVYNTI 843
Db      600 PKRPAKGSAPINLSRLNOVLRMALGSVEILSRHCPIMWYGNGLKFLERAVYNTI 659
Oy      844 YPITSIPILMYCTLLAVCLFTNOFIPOISNIAIMFLSLFISFATGILEMRSGVID 903
Db      660 YPITSIPILMYCTLLAVCLFTNOFIPOISNIAIMFLSLFISFATGILEMRSGVID 719
Oy      904 EWMRNOQFVIGVSAHLFAVFOGILKVLADITNFTYKSKASDEGDAELYLFKWTTL 963
Db      720 EWMRNOQFVIGVSAHLFAVFOGILKVLADITNFTYKSKASDEGDAELYLFKWTTL 779
Oy      964 LIPPTLLILNVGVAVGSAHNSGYSOMPLGLFFAFWYVHLHYPLKGLMGROR 1023
Db      780 LIPPTLLILNVGVAVGSAHNSGYSOMPLGLFFAFWYVHLHYPLKGLMGROR 839
Oy      1024 PPTIVVWMAILLASIFSLMVRIDPPTSRVTGPDILGCGINC 1065
Db      840 PPTIVVWMAILLASIFSLMVRIDPPTSRVTGPDILGCGINC 881

```

```

RESULT 2
US-08-960-048-6
; Sequence 6, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Gossypim hirsutum
US-08-960-048-6

```

```

Query Match      61.2%; Score 3473.5; DB 4; Length 974;
Best Local Similarity 62.1%; Pred. No. 0;
Matches 660; Conservative 129; Mismatches 176; Indels 97; Gaps 15;

```

```

Oy      12 MNIVPOTQICSDNYGKTVDGDFVACDICSFPVCPCEYERKGNOSCPQCKTRYKR 71
Db      2 MESSGVV-YCHTCEHGVGLNVGEBEPFVACHCNPNPKSCFEYDLKGRACLRG----- 54
Oy      72 LKGSRAIPGDKDEGLADGCTVEFENPQKEKISERMLGHLRKGKEEMGEPOYKREVSH 131
Db      55 --GSP-----YDENLDD-----VERATGDSYMAHL--NKSQDVG-----IHA 90
Oy      132 NMLPRLTSRODTSSEFSASPERLSVSTIAGKRLPYSSDVNQSNNRIYDPVGLGNVA 191
Db      91 RIISSVSTIJDSEMAEDN-----GNSI 111

```

```

Oy      192 WKERVDMKMKQKCKNTGVPSTQASERGVGVIDASTDILADEALLND---EAROPLSRK 247
Db      112 MKRVESEMKKKKKKKKPAFTKVERAE-----IPPEQOMEDKKPAPDASQPLSTI 161
Oy      248 VSIPISSRINPYRNVIMLRVILCLFLHYITNPVPNAPFALMLVSYICEIMFALSWLDOF 307
Db      162 IPIPKSRILAPYRIVIIIMRLIILGLFPHYVNTNVDASFGIMLTISVCEIMFASWLDOP 221
Oy      308 PKWFPVNRREYLDRLARVDRREGESQLAVIDFVSTVPLKEPPLVNTANTVLSILAVY 367
Db      222 PKWFPVNRREYLDRLARVDRREGESQLAVIDFVSTVPLKEPPLVNTANTVLSILAVY 281
Oy      368 PVDKVCYVFDGGAAMLSESLAETSEFARKVVPFCCKYSIEPRADWYFAAKIDYLDK 427
Db      282 PVDKVCYVFDGGAAMLSESLAETSEFARKVVPFCCKYSIEPRADWYFAAKIDYLDK 341
Oy      428 VOTSFYKDRRAKREIEEKKIRINALVSALAKCEGGMWMOGTPPMPGNTGDHPMIOY 487
Db      342 VOPSFYKERRAKRDEEYKIRINALVAAOKTPDGMWMOGTPPMPGNTGDHPMIOY 401
Oy      488 FLGONGGLDAEGNELPRLVYVSREKRPGRQHHKKAGAMNALVSAVLNNGPILMLDGD 547
Db      402 FLGONGGLDAEGNELPRLVYVSREKRPGRQHHKKAGAMNALVSAVLNNGPILMLDGD 461
Oy      548 HYINNSKALREAMCFLLMDPNLGRSVYVOPQRFQDIGNDRYANNNTVFEDINR 607
Db      462 HYINNSKALREAMCFLLMDPNLGRSVYVOPQRFQDIGNDRYANNNTVFEDINR 521
Oy      608 IGPVYVGTGCVFNRNALGYEPPIVYKRRKSLSLKCGSRKNSKKKSDKK 667
Db      522 IGPVYVGTGCVFNRNALGYEPPIVYKRRKSLSLKCGSRKNSKKKSDKK 579
Oy      668 HDSTVTPVFNLDIEGVEGAGFDD--EKALIMSOMSLKRRFGOSAVFVASTLMENG 726
Db      580 HDSTVTPVFNLDIEGVEGAGFDD--EKALIMSOMSLKRRFGOSAVFVASTLMENG 539
Oy      727 SATPENILKEAHIHIVISCGYEDKSDMGMEIGWITGVTEDILTFKMHARGWSITYCM 786
Db      634 SATPENILKEAHIHIVISCGYEDKSDMGMEIGWITGVTEDILTFKMHARGWSITYCM 693
Oy      787 PAFKGSAPINLSRLNOVLRMALGSVEILSRHCPIMWYGNGLKFLERAVYNTI 845
Db      694 PAFKGSAPINLSRLNOVLRMALGSVEILSRHCPIMWYGNGLKFLERAVYNTI 753
Oy      846 ITSIPILMYCTLLAVCLFTNOFIPOISNIAIMFLSLFISFATGILEMRSGVID 905
Db      754 ITSIPILMYCTLLAVCLFTNOFIPOISNIAIMFLSLFISFATGILEMRSGVID 813
Oy      906 WRNEQFVIGVSAHLFAVFOGILKVLADITNFTYKSKASDEGDAELYLFKWTTL 965
Db      814 WRNEQFVIGVSAHLFAVFOGILKVLADITNFTYKSKASDEGDAELYLFKWTTL 872
Oy      966 PPTLLILNVGVAVGSAHNSGYSOMPLGLFFAFWYVHLHYPLKGLMGROR 1025
Db      873 PPTLLILNVGVAVGSAHNSGYSOMPLGLFFAFWYVHLHYPLKGLMGROR 932
Oy      1026 TIVVWMAILLASIFSLMVRIDPPTSRVTGPDILGCGINC 1065
Db      933 TIVVWMAILLASIFSLMVRIDPPTSRVTGPDILGCGINC 974

```

```

RESULT 3
US-08-960-048-7
; Sequence 7, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987

```

; PRIOR FILING DATE: 1996-10-29  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 685  
 ; TYPE: PRF  
 ; ORGANISM: Gossypium hirsutum  
 US-08-960-048-7

Query Match 49.4%; Score 2803; DB 4; Length 685;  
 Best Local Similarity 74.5%; Pred. No. 5.9e-276;  
 Matches 512; Conservative 77; Mismatches 76; Indels 22; Gaps 6;

```

QY 396 ARKWPFCKKXISIPAPWYFAKIDYLDKDYOTSPFKDRRAKREYEERKRNALVVS 455
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 ARWVPFCKKHNVPRAPPEFYNEKIDYLDKDYOTSPFKDRRAKREYEERKRNALVVA 60
QY 456 KALKCPBGMVWMDGTWPGNNTGDHPGMIOVFLQNGGLDAEGNELPLVYVSREKRG 515
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 KAKCPBGMVWMDGTWPGNNTGDHPGMIOVFLQNGGLDAEGNELPLVYVSREKRG 120
QY 516 FOHKKAGAMALYRVSANVLNPGPILNLDGHTINNSKALREACFLMDPRLKQVYV 575
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 YOHKKAGAMALYRVSANVLNPGPILNLDGHTINNSKALREACFLMDPRLKQVYV 180
QY 576 QPQPFQIDKNDYRANRNVFEDINLGLDIOGPPYVGVGVNFRALGYEPPIKVK 635
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 QPQPFQIDKNDYRANRNVFEDINLGLDIOGPPYVGVGVNFRALGYEPPIKVK 240
QY 636 HKK-----PSLSKLCGSGRRKSKAKKE-----SDKKSGRH--TDSTVPFN 677
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 RPKWTCDCMPSCGCCGSRKSKKKEKGLGLGLLGYKKKKMMGNVYKKSAPVFD 300
QY 678 LDOIIEBEGAGFED-EKALLMSQMSLEKRGQSAVPASTLMENGVPSPATPENLKE 736
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 LEEIEEGL--GYELEKSTLMSQNEKRGQSPVFASTLMENGSLPEGTNSTSLIKE 358
QY 737 AIHYISGIEDKSDMGMEIGWYGVTEEDILTGFKMARHSYICMPLKAPFGSAPIN 796
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 AIHYISGIEDKSDMGMEIGWYGVTEEDILTGFKMARHSYICMPLKAPFGSAPIN 418
QY 797 LSDRLNOVLRWALGSVELFSRHCPWYNGYGRLEFERAYVNTTYPITSIPLMYCT 856
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 LSDRLNOVLRWALGSVELFSRHCPWYNGYGRLEFERAYVNTTYPITSIPLMYCT 478
QY 857 LLAVCLTNPOTIIOISINISWFLSLFATGILERMWSGVGIDEMWNEQFWYIG 916
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479-IPAVCLTNGKFIPTLNLTSVWFLALFLSIATGVLELRMSGVSIODMWRNEQFWYIG 538
QY 917 VSAHLFAVFOGILVLAGIDTNPFTVASKASDEDDPAELVFKMTLLIPTLLILVLY 976
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 VSAHLFAVFOGILVLAGIDTNPFTVASKAD-DTEFEGELVFKMTLLIPTLLILVLY 597
QY 977 GVAVGSYAINSGYOSGMPLEGLLFFAFWVIVHLVPELKLGMGRNRTPTIVVSVLLA 1036
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 598 GVAVGSYAINSGYOSGMPLEGLLFFAFWVIVHLVPELKLGMGRNRTPTIVVSVLLA 657
QY 1037 SIFSLWVRIDPFTSRVTGPDIIECGI 1063
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 658 SIFSLWVRIDPFTSRVTGPDIIECGI 684
  
```

RESULT 4  
 US-08-960-048-11  
 ; Sequence 11, Application US/08960048C  
 ; Patent No. 6271443  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stalder, D. et al.  
 ; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter  
 ; FILE REFERENCE: 15621/01/US  
 ; CURRENT FILING DATE: 1997-10-29

; PRIOR APPLICATION NUMBER: 60/029,987  
 ; PRIOR FILING DATE: 1996-10-29  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 11  
 ; LENGTH: 693  
 ; TYPE: PRF  
 ; ORGANISM: Escherichia coli  
 US-08-960-048-11

Query Match 4.9%; Score 278; DB 4; Length 693;  
 Best Local Similarity 19.3%; Pred. No. 7.7e-19;  
 Matches 153; Conservative 86; Mismatches 191; Indels 362; Gaps 30;

```

QY 260 WVMIRLVILCLFLHYRT-----NPNPNAFLMVLVSVICELPALSMLIDQPKMPV 313
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 21 MLIVLSTVSCRITWRTSTLNMDDPVSIVGLLILFLAITVAVIL--VLGFQVVPPL 78
QY 314 NRETYLDRLARVDEGEPSOLAADVIFSTVDPLEBPPLVANTVLSILAVDPVDRVS 373
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 79 NRQP-----VPLPKDMSLMP-----VDIFVPTYN--EDLVNKKNTIYASLGDMPKDKLN 127
QY 374 CYVEDDGAMLSFESLAESEFARKWVPCKKYSIEPRAPWYFAKIDYLDKDYOTSFV 433
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 128 IMILDDG----- 135
QY 434 KDRAMKREYEERKRNALVSKALCPBGMVWMDGTWPGNNTGDHPGMIOVFLQNG 493
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 -----REFRQF----- 146
QY 494 GLDAEGNELPLVYVSREKRGPOGHHKKAGAMALYRVSANVLNPGPILNLDGHTINNS 553
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 GVK-----YAKRT--HEHAKAGINNALAKA-----KEFVSIDCDHVPTRS 188
QY 554 KALREACFLMDPRLKQVYVQPFQRE--DGIDKN--DRYANRNTVEFDINLRLGLD 606
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 189 FLOWMTGWFLKE--KOLAMQTPHHFSPDPFERNLGRFRKTPNEGTLFGLVQDGN 244
QY 607 GIQGVYVGTGCVNFRKTLVLYGEYPIKVKHKKPSLSKLCGSGRRKSKAKKESKSG 666
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 MMDATFFGSCAVIR-- 262
QY 667 RHTDSTVVFVNDLIEBEGAGFDEKALLMSQMSLEKRGQSAVPASTLMENGVP 726
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 263 -----LDEL----- 269
QY 727 SATPENLKEAIVHISGIEDKSDMGMEIGWYGVTEEDILTGFKMARHSYICMPLK 786
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 270 -----AVE-----TVDATHTSLRLHRGYSAYM--RI 296
QY 787 PAFKGSAPINLSDRLNOVLRWALGSVELFSRHCPWYNGYGR-LKLEFERAYVNTTYP 845
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 297 POAGLATVESLASHIGORLIRWARGVOI-FRLDNL--TGKGLKFAORLCYVAMHFE 351
QY 846 ITSISLWYCT-----LLAVCLTNPOTIIOISINISWFLSLFATGILERMWSGV 886
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 352 LSGIPRLFLTPALPAFLLLHAYIYAPALMALAFVLPMMHINSLNSIQGRKRSPSE 411
QY 887 IFAT-----GILERMWSGVGIDEMWNEQF--WVIGVSAHLFAV 925
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 412 IYETVLAWYIAPPTLVALINPHKGFNTYAKGGGLVE--EEYDVWYIS--RPIYIFVL 465
QY 926 OGILVLAGIDTNPFTVASKASDEDDPAELVFKMTLLIPTLLILVLY 976
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 466 LNLGVAVGI-----WRYFGPTTEMLTVVSNWVYFNL 501
QY 977 GVAVGSYAINSG 988
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 502 VLGGAVANVS 513
  
```

RESULT 5  
 US-08-960-048-12









DB 453 ASKINKGR-YSEFSEVETTMALFLVRYTVTLTSPRGKFNVDKGLLEKGYFDL- 509  
QY 958 FKMTLLIPTTLVLVNVAGVYVAINSGYOSWGLGGLFFAFVYVYHLPFLKGL 1017  
DB 510 ---GAVYPNILLGLIMEGLARGV-YELSPGHLDQIAERAYLLNSAWAMLSLIIILAI 564  
QY 1018 -MGRO-----NRTPTIVV 1030  
DB 565 AVGRETOOKRNSHRIPATIPV 585

RESULT 10  
PCT-US91-01726-4  
Sequence 4, Application PC/TUS9101726  
GENERAL INFORMATION:  
APPLICANT: Brown, Malcolm R  
APPLICANT: Saxena, Indar M  
APPLICANT: Lin, Fong C  
TITLE OF INVENTION: Recombinant Cellulose Synthase  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David L. Parker  
STREET: 600 Congress Avenue Suite 2300  
CITY: Austin  
STATE: Texas  
COUNTRY: United States  
ZIP: 78701

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01726  
FILING DATE: 19910314  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 494093  
FILING DATE: 15-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker Mr, David L  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512-320-7200  
TELEFAX: 512-474-7577  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 723 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US91-01726-4

Query Match 3.5%; Score 196; DB 5; Length 723;  
Best Local Similarity 20.1%; Pred. No. 1.9e-10;  
Matches 161; Conservative 93; Mismatches 241; Indels 306; Gaps 36;

QY 251 PSSRNPRMYIMLFLVLCFLAYRITNPVNAFALM-----VSVICEIMWLSWTL 304  
DB 70 PSRSQILF-LEVLSTGLVSL-RYLWRLTETL--SFDTWLOGGLGMLLVAELIYALMMLFL 125  
QY 305 DQFPWPFVNRETYIDRLALNRDRGCEPSQLAANDIFSTVDPLKEPLVYTNVYLSILA 364  
DB 126 SYFQTIADLPHRP---LPL---PPNDDEMTVDIFVPTYN---EELSTVRLTVLGSIG 174  
QY 365 VDYPVKDYCVYDGAAMLSPESLAETSEFARKWVPCKKYSIEPRAPEWYFAAKIDYL 424  
DB 175 IDMPREKRVHILDDG-----RRPE--FAA----- 197  
QY 425 KDKVOTSYVKDRAMKREYEERKINIMALVSKALKCPREBGMVMDGTPMPGNNGTDDHPCM 484  
DB 198 -----FAAEC----- 202

QY 485 IOVFLGONGGLDAEGNELPRLVYVSRERKRGPHGHHKAGAMNALRVSAVLNCPILNL 544  
DB 203 ---GAN-----YIARPT-----NEAAKGNLNYAIG---HTDGYILLF 235  
QY 545 DCDHYINNSKALREACMLM-DPNLKOVCYVOPQREDGIDKNDRYANRNTVFEDINLR 603  
DB 236 DCDH-VPTRAFLQITGMWVEDP-----KIALMQTPHHFSPDP----- 273  
QY 604 GLDGIQGVYVVGCGVFNRTALXGEPRPIKVKHKKRPSLSKLGSGSRKKNSAKKESDOK 663  
DB 274 -----FORNLSAGYRTP----- 285  
QY 664 KSGRHTDSTVPVFNIDIEEGVAGFDEKALLMSOMSLERKFGQSAVFVSTLMENG 723  
DB 286 -----PEGNL---FYGVYQDND-----FNDATFFGSCAI----- 313  
QY 724 VPPSATPENLKEAIVHISCGYEDKSDMGEITGWIYGSYTEDITLGFKNHARGMSIYCM 783  
DB 314 -----LRTTAIEQIG-----GFATQVTEBAHTALKMORLGWSTAYL- 350  
QY 784 PKLPFKSAPRLNLSDRNLQVLRNALGSVEILFSRHCPTWYGNR-LKFLERFAYVNT 842  
DB 351 -RIPLAGGLATERLLHIGQVRMARGLQI-FRIDNPLE---GRGLSWGRLCYLSM 404  
QY 843 IYPTISIPLMYCTLLAVCLFTNPFIPQISNIASIMFLSLPSIFATGILEMRSGVG- 901  
DB 405 TSFLFAVRVIRFLSSPLAFLEFGQNIITASP-----LALLAVIRPM-FHAYGT 452  
QY 902 ---IDEMWRNEQFW-VIGVSAHLFAVFGILKVLADITNFTVTSKASDEGDPAEYVL 957  
DB 453 ASKINKGR-YSEFSEVETTMALFLVRYTVTLTSPRGKFNVDKGLLEKGYFDL- 509  
QY 958 FKMTLLIPTTLVLVNVAGVYVAINSGYOSWGLGGLFFAFVYVYHLPFLKGL 1017  
DB 510 ---GAVYPNILLGLIMEGLARGV-YELSPGHLDQIAERAYLLNSAWAMLSLIIILAI 564  
QY 1018 -MGRO-----NRTPTIVV 1030  
DB 565 AVGRETOOKRNSHRIPATIPV 585

RESULT 11  
US-08-119-773-4  
Sequence 4, Application US/08119773  
Patent No. 5460942  
GENERAL INFORMATION:  
APPLICANT: Chou, Janice Y.  
APPLICANT: Lei, Ke-Jian  
APPLICANT: Shelly, Leslie L.  
TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourile and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/119,773  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Kenneth A.  
REGISTRATION NUMBER: 31,677  
REFERENCE/DOCKET NUMBER: 15280-175  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 357 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..357  
OTHER INFORMATION: /note= "The G-6-Pase amino acid sequence mutated  
OTHER INFORMATION: from Arg to Cys at position 83."  
US-08-119-773-4

Query Match 2.2%; Score 124; DB 1; Length 357;  
Best Local Similarity 19.8%; Pred. No. 0.0011;  
Matches 79; Conservative 44; Mismatches 122; Indels 154; Gaps 16;

QY 736 EAIHWISCGYEDKSDWMEIGWYGSVTEEDILTFGRMARGRSICYMKELPAFKGSAPI 795  
DB 14 GSTHYLVNYQDSQD-----FLVSVIADLNNAFVLPPIW---FHLQEAVGIKLLMVA 65  
QY 796 NISDRNLNOLRNALSGVELLFSRHCPHWG----- 825  
DB 66 VIGDWLNLVEKW-----LTFGQ-CPYWVLDTDYYSNTSVPLIKOPVTCETGPGSPSG 118  
QY 826 -----YNGRLKFLERFAYVNTTIYPTSIPILMYCTLLAVCL----- 862  
DB 119 HAMGTAGYVYVAVTSTLSIFGCKIKPTYFRCLNVIW-----LGFMAVOLNVCISRTY 172  
QY 863 ----FTNOFIPOISNIA-----SIWFLSL-----FLSIFATGILEMRSGVG 901  
DB 173 LAHPHQYVAVGLSGIAVTEFTHSHISYNAKLKYPFLTFPFSFAIGFY-LTLKGIG 231  
QY 902 IDEWWRNE--QFWVIGVSAHLFAVFGILKVLADITNFTYTSKASDEDDGFAELYLEK 959  
DB 232 VDLMLTLEKQKRCQPEKWH-----IDTT----- 256  
QY 960 WTTLLIPTTLILVNLVGVAVSYAINSG-----YOSMGPLFGKLPFAFWIYHL 1010  
DB 257 -----PPASLLKLN-GTLFGLGLALNSMYSRESCKKLSKWLPPCLSIYASIVLHLV 308  
QY 1011 YPFLKGLMGRNRPPIVYVMSVLLASIFSLMVRIDPF 1049  
DB 309 FDSLK-----PPSQVELVEFVLSFCKSAVVPPLASVSIVPY 343

RESULT 12  
US-08-119-773-6

; Sequence 6, Application US/08119773  
; Patent No. 5460942  
; GENERAL INFORMATION:  
; APPLICANT: Chou, Janice Y.  
; APPLICANT: Lei, Ke-Jian  
; APPLICANT: Shelly, Leslie L.  
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Stewart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/119,773  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Kenneth A.  
REGISTRATION NUMBER: 31,677  
REFERENCE/DOCKET NUMBER: 15280-175  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 357 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..357  
OTHER INFORMATION: /note= "G-6-Pase mutated at position 295 with Arg  
OTHER INFORMATION: changed to Cys"  
US-08-119-773-6

Query Match 2.0%; Score 116; DB 1; Length 357;  
Best Local Similarity 19.5%; Pred. No. 0.0073;  
Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;

QY 736 EAIHWISCGYEDKSDWMEIGWYGSVTEEDILTFGRMARGRSICYMKELPAFKGSAPI 795  
DB 14 GSTHYLVNYQDSQD-----FLVSVIADLNNAFVLPPIW---FHLQEAVGIKLLMVA 65  
QY 796 NISDRNLNOLRNALSGVELLFSRHCPHWG----- 825  
DB 66 VIGDWLNLVEKW-----LTFGQ-CPYWVLDTDYYSNTSVPLIKOPVTCETGPGSPSG 118  
QY 826 -----YNGRLKFLERFAYVNTTIYPTSIPILMYCTLLAVCL----- 862  
DB 119 HAMGTAGYVYVAVTSTLSIFGCKIKPTYFRCLNVIW-----LGFMAVOLNVCISRTY 172  
QY 863 ----FTNOFIPOISNIA-----SIWFLSL-----FLSIFATGILEMRSGVG 901  
DB 173 LAHPHQYVAVGLSGIAVTEFTHSHISYNAKLKYPFLTFPFSFAIGFY-LTLKGIG 231  
QY 902 IDEWWRNE--QFWVIGVSAHLFAVFGILKVLADITNFTYTSKASDEDDGFAELYLEK 959  
DB 232 VDLMLTLEKQKRCQPEKWH-----IDTT----- 256  
QY 960 WTTLLIPTTLILVNLVGVAVSYAINSG-----YOSMGPLFGKLPFAFWIYHL 1010  
DB 257 -----PPASLLKLN-GTLFGLGLALNSMYSRESCKKLSKWLPPCLSIYASIVLHLV 308  
QY 1011 YPFLKGLMGRNRPPIVYVMSVLLASIFSLMVRIDPF 1049  
DB 309 FDSLK-----PPSQVELVEFVLSFCKSAVVPPLASVSIVPY 343

RESULT 13

US-08-119-773-5  
; Sequence 5, Application US/08119773  
; Patent No. 5460942  
; GENERAL INFORMATION:  
; APPLICANT: Chou, Janice Y.  
; APPLICANT: Lei, Ke-Jian  
; APPLICANT: Shelly, Leslie L.

;; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND  
;; TITLE OF INVENTION: PROTEIN AND RELATED MUTATIONS  
;; NUMBER OF SEQUENCES: 36  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend Hourie and Crew  
;; STREET: Steuart Street Tower, One Market Plaza  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: US  
;; ZIP: 94105-1493  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/119,773  
;; FILING DATE: 10-SEP-1993  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Weber, Kenneth A.  
;; REGISTRATION NUMBER: 31,677  
;; REFERENCE/DOCKET NUMBER: 15280-175  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 543-9600  
;; TELEFAX: (415) 543-5043  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 346 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHEICAL: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 1..346  
;; OTHER INFORMATION: /label= Truncated prot  
;; OTHER INFORMATION: /note= \*G-6-pase truncated due to a C to T  
;; OTHER INFORMATION: mutation at nucleotide 1118 converting a glutamine  
;; OTHER INFORMATION: to a stop codon at 347\*  
US-08-119-773-5

Query Match 2.0%; Score 115; DB 1; Length 346;  
Best Local Similarity 19.5%; Pred. No. 0.0087;  
Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;

OY 736 EAIHVISCYEDKSDMGMEIGYGVYEDILITGFKHARGRSIYCPKLPAPFGSAPI 795  
DB 14 OSTRHLYQVNYDSQDM-----FILVSVIADLRNAPYVLPPIW---FHLQEAVGIKLLMVA 65  
OY 796 NLSDRNLQVLRWALGSVELFSRHCPYWG----- 825  
DB 66 VIGDWLNLVFKW-----ILFGQR-PYMWVLDTDYYSNTSVPLIKOPVTCETGPGSPG 118  
OY 826 -----YNGRLKFLERFAVNTTITPITSIPLMYCTLLAVCL----- 862  
DB 119 HAMGTAGYVYVWVTSTLSIFQGIKIPYRFRCLNVLW-----LGFNAVOLNVLCLSRIV 172  
OY 863 -----FTNOFIIPQISNIA-----SIFPLSL-----FLSIFATGILEMRWSGVG 901  
DB 173 LAHFPHQVAVGVSIGIAVTETFSIHISIIYNSLKKYFLITFFLFSFAIGFY-LLLKGLG 231  
OY 902 IDEWWRNE--QFWVIGVSAHLFAVFOGLKVLAGIDINFTVTSKASDEDDGFALYLFK 959  
DB 232 VDLMLTLEKARWCEQPEMVH-----IDTT----- 256  
OY 960 WTTLLIPPTLLIYLVGVVAVGSVAINSG-----YOSWGPLFGKLPFAFWYVYHL 1010  
DB 257 -----PFASLKLNKLTGLFGLALNNSMWRSCGKLSKWLFPRLSIVASIVLVLHV 308

OY 1011 YPFLKGLMGKRONRPTIYVWWSVLLASIFSLLWRIDPF 1049  
DB 309 FDSLK-----PPSQVELVIFYVLFSCKSAVAVPLASVIVY 343

RESULT 14  
US-08-119-773-2  
; Sequence 2, Application US/08119773  
; Patent No. 5460942  
; GENERAL INFORMATION:  
; APPLICANT: Chou, Janice Y.  
; APPLICANT: Lei, Ke-jian  
; APPLICANT: Shelly, Leslie L.  
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATASE: THE GENE AND  
; TITLE OF INVENTION: PROTEIN AND RELATED MUTATIONS  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Hourie and Crew  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/119,773  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Kenneth A.  
; REGISTRATION NUMBER: 31,677  
; REFERENCE/DOCKET NUMBER: 15280-175  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-5043  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 357 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-119-773-2

Query Match 2.0%; Score 115; DB 1; Length 357;  
Best Local Similarity 19.5%; Pred. No. 0.0092;  
Matches 78; Conservative 44; Mismatches 123; Indels 154; Gaps 16;

OY 736 EAIHVISCYEDKSDMGMEIGYGVYEDILITGFKHARGRSIYCPKLPAPFGSAPI 795  
DB 14 OSTRHLYQVNYDSQDM-----FILVSVIADLRNAPYVLPPIW---FHLQEAVGIKLLMVA 65  
OY 796 NLSDRNLQVLRWALGSVELFSRHCPYWG----- 825  
DB 66 VIGDWLNLVFKW-----ILFGQR-PYMWVLDTDYYSNTSVPLIKOPVTCETGPGSPG 118  
OY 826 -----YNGRLKFLERFAVNTTITPITSIPLMYCTLLAVCL----- 862  
DB 119 HAMGTAGYVYVWVTSTLSIFQGIKIPYRFRCLNVLW-----LGFNAVOLNVLCLSRIV 172  
OY 863 -----FTNOFIIPQISNIA-----SIFPLSL-----FLSIFATGILEMRWSGVG 901  
DB 173 LAHFPHQVAVGVSIGIAVTETFSIHISIIYNSLKKYFLITFFLFSFAIGFY-LLLKGLG 231  
OY 902 IDEWWRNE--QFWVIGVSAHLFAVFOGLKVLAGIDINFTVTSKASDEDDGFALYLFK 959  
DB 232 VDLMLTLEKARWCEQPEMVH-----IDTT----- 256  
OY 960 WTTLLIPPTLLIYLVGVVAVGSVAINSG-----YOSWGPLFGKLPFAFWYVYHL 1010

Db 257 -----PRASLIKLN-GTLFGILALNNSMYRESCGKLSKWLPERLSIVASVLLHV 308  
 QY 1011 YPELKGIMGRONTPTIVVWVSIVLASISFELMWRIIDPF 1049  
 Db 309 FDSLK-----PPSQVELFVYLSFCKSAVPLASVYIPY 343

RESULT 15  
 US-09-000-094-46  
 : Sequence 46, Application US/09000094  
 : Patent No. 6365160

GENERAL INFORMATION:

APPLICANT: WEBB, Elizabeth Ann  
 MARGETS, Mary Bridgid  
 COX, John Cooper

MCWILLAN, Nigel Alan John  
 WILLIAMS, Mark Phillip  
 MOLONEY, Margaret Bridget  
 Holland

EDWARDS, Stirling John  
 TITLE OF INVENTION: PAPILLOMAVIRUS POLYPEPTIDE CONSTRUCTS

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER  
 STREET: 3000 K Street, N.W.  
 CITY: Washington

STATE: D.C.  
 COUNTRY: U.S.A.

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/000,094

FILING DATE: 21-Apr-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/AU96/00473

FILING DATE: 26-JUL-1996

APPLICATION NUMBER: AU PN 4439/95

FILING DATE: 27-JUL-1995

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 017227/0137

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 1587 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 46:

US-09-000-094-46

Query Match 1.98; Score 110; DB 4; Length 1587;  
 Best Local Similarity 19.38; Pred. No. 0.44;  
 Matches 114; Conservative 77; Mismatches 221; Indels 180; Gaps 30;  
 QY 2 ESEGETAGKPMK---NIVPOTGICSDNGKTVGDRFVACDICSFPVCRPCVEYERKDC 58  
 Db 877 EVDGDS-OPLKQHOIVTCCG-CDSNRLVYQ-----CTETIREY 917  
 QY 59 NQ-----SCPOCKTRYKRLKSPADGKDEGLADGTVGFVNYPOKEKISERMLGW 110  
 Db 918 OQLLGLTINIVPIC-----APKTSMDDDSTENGSGCTGMFVNAIVQHPGT 968  
 QY 111 HLTRGKGEMGEPOY-----DKEVSHNHL--PLTTSRODTSGEFSASPERLSVSSST 161

Db 969 QISDDEDEVEDSGYMDVDFIDDSNTHNSLEAQLFNQEAQDTH-----ATY 1017  
 QY 162 AGGR-----LPYSSDNO-----SPRRRTVDVVGNGVAMK-----ERV-D-G 198  
 Db 1018 QDLKRRYLGSPVSPINTIAEAVESEISPR--LDAIKILTRQPKYKRLFOETRELDSG 1074  
 QY 199 WKMKQ-EKNTG-PVSTQAAASERG--GVDDIDASTDILADBALLNDEARQPLSRKVSIPSSR 254  
 Db 1075 YGSEVEAGTGTVEKHGVPENGDGDQENDGTGRDIEGE-----HTEAEAPTNS 1123  
 QY 255 INPYRMVIMRLVILCLFLHYRTNPVNAFALMYSVI-----CEYFALSM--- 302  
 Db 1124 VREHAGTAGITLLEIKCKDLRAALLGFKKECFGLSFIDLIRPKSDKTTCLDWNVAAGFGIH 1183  
 QY 303 --LLDQPKWF-PVNRETYLDRLA-----LRYDEGEPSOLA---AVDIFVSTVD 346  
 Db 1184 HSTSEAFQKILIEPLSYAHQWLTNMGVLLVLLFKVKNKSSTVAKRLATLTLINIPENQ 1243  
 QY 347 PLKEPPLV-----TANTVLSILAADVDPVKVSCYVDDGAAMLSFE---- 387  
 Db 1244 MLEPPKIQSGVALYWPFTGISNASTVIG---EAPENITROTVEHGLADSGFKLTETM 1299  
 QY 388 -----SLAETSEFARKVVPCKKYSIEPRAPEMTFPAKIDYLRKQVOTSPYKDRRAK 440  
 Db 1300 VQWAYDNDICESEISA---FEYAQRGDPDFSNARA-----FLNSNMQAKYVKDCATMC 1348  
 QY 441 REYEERKIRNMLVSKALCKPREGVWMOGTTPAGNNTGDHGMIOVFIGON 492  
 Db 1349 RHYKHAEMR-----KMSIKQWIKHRSKIEG--TGMMKPIVOFLRHON 1389

Search completed: June 16, 2003, 11:08:57  
 Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:01:28 ; Search time 45 Seconds  
(without alignments)  
3153.598 Million cell updates/sec

Title: AAC39336  
Perfect score: 5677  
Sequence: 1 MESEGETAGPKPMKNIVPQTC.....IDPFTSRVTGPDIECGINC 1065

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_101002:\*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5655	99.6	1065	19	AAW33819
2	5654	99.6	1065	21	AAAG48733
3	5654	99.6	1065	23	ABB93356
4	5597	98.6	1054	21	AAAG48734
5	5066	89.2	959	21	AAAG48735
6	4572	80.5	1077	21	AAAY84108
7	4572	80.5	1077	21	AAAY84114
8	4564	80.4	1079	23	AAU10496
9	4529.5	79.8	1076	21	AAAY84121
10	4418	77.8	1043	21	AAAY84119

11	4028.5	71.0	1075	21	AAAY84109	Amino acid sequenc
12	4028.5	71.0	1075	21	AAAY84115	Amino acid sequenc
13	4028.5	71.0	1075	21	AAAY84120	Amino acid sequenc
14	4024	70.9	1074	21	AAAY84110	Amino acid sequenc
15	4024	70.9	1074	21	AAAY84117	Amino acid sequenc
16	4024	70.9	1074	21	AAAY84112	Amino acid sequenc
17	3984	70.2	1081	19	AAW33817	Herbicideidally activ
18	3984	70.2	1081	23	ABB93199	Herbicideidally activ
19	3974	70.0	1081	19	AAW33820	Arabidopsis cellu
20	3974	70.0	1081	19	AAW33820	Arabidopsis cellu
21	3900	68.7	1058	21	AAAY84107	Corn cellulose syn
22	3875.5	68.3	1086	21	AAAY84111	Amino acid sequenc
23	3875.5	68.3	1086	21	AAAY84116	Amino acid sequenc
24	3875.5	68.3	1086	21	AAAY84113	Corn cellulose syn
25	3843.5	67.7	1094	21	AAAY84118	Amino acid sequenc
26	3843.5	67.7	1094	21	AAAY84118	Amino acid sequenc
27	3843.5	67.7	1165	23	AAAY58834	Corn cellulose syn
28	3811	67.1	1026	23	ABB93522	Herbicideidally activ
29	3804.5	67.0	1039	21	AAAY58837	Soybean cellulose
30	3801	67.0	1065	23	ABB91880	Herbicideidally activ
31	3748.5	66.0	1039	19	AAW73309	Cellulose synthase
32	3741	65.9	1043	23	ABB93701	Herbicideidally activ
33	3670	64.6	1069	23	ABB93411	Herbicideidally activ
34	3647.5	64.3	1084	23	ABB93949	Herbicideidally activ
35	3628	63.9	1084	23	ABB93304	Herbicideidally activ
36	3596	63.3	1080	22	AAW37895	Arabidopsis thalia
37	3596	63.3	1084	19	AAW33818	Arabidopsis cellu
38	3590	63.2	1088	23	ABB91831	Herbicideidally activ
39	3522	62.0	821	23	AAU10495	Corn cellulose syn
40	3476.5	61.2	974	19	AAW73308	Cellulose synthase
41	3473.5	61.2	974	19	AAW60039	Cotton cellulose s
42	3348	59.0	955	21	AAW16338	Eucalyptus grandis
43	3317.5	58.4	958	23	ABB92881	Herbicideidally activ
44	3216	56.6	793	21	AAAY58839	Soybean cellulose
45	3143	55.4	946	22	AAW37893	Populus tremuloid

#### ALIGNMENTS

RESULT 1	AAW33819	standard; Protein; 1065 AA.
ID	AAW33819	
XX	AAW33819;	
AC		
XX	06-JUL-1998 (first entry)	
DT		
XX	Arabidopsis cellulose biosynthetic protein Ath-B.	
DE		
XX	Cellulose; cellulose synthase; RSW1 protein; beta-1,4-glucan;	
KW	transgenic plant.	
XX		
OS	Arabidopsis thaliana var. Columbia.	
XX		
PN	W09800549-A1.	
PD	08-JAN-1998.	
XX		
PF	24-JUN-1997; 97WO-AU00402.	
XX		
PR	27-JUN-1996; 96AU-0000699.	
XX		
PA	(CSIR ) COMMONWEALTH SCI & IND RES ORG.	
PA	(AUSU ) UNIV AUSTRALIAN NAT.	
PI	Artoli A, Betzner AS, Peng L, Williamson RE;	
DR	WPI: 1998-086974/08.	
DR	N-PSDB: AAU06567.	
XX		
PT	DNA encoding cellulose biosynthetic enzyme - useful for manipulation	
PT	of cellulose and beta-1,4-glucan	

XX Claim 29, page 144-149; 207pp; English.  
XX  
XX This polypeptide is encoded by cDNA clone Ath-B (see AAV06567)  
CC that was isolated from an Arabidopsis thaliana cDNA library using  
CC PCR primers (see AAT99632-34) based on cellulose synthase RSW1  
CC genomic clone 23H12 (see AAV06563) and EST clone AAT20782 (see  
CC AAV06562). It is closely related to Arabidopsis cellulose synthase RSW1  
CC protein (see AAW3816-17). Claimed nucleic acid molecules (see  
CC AAV0562-69) coding for claimed polypeptides (see AAW3816-20 and  
CC AAW46202) involved in cellulose biosynthesis can be used to  
CC manipulate the cellulose and/or beta-glucan content of transgenic  
CC plants. Expression of nucleic acids in the sense orientation  
CC increases the level of cellulose and reduces the level of  
CC non-crystalline beta-1,4-glucan and starch, providing plants with  
CC modified strength and/or shape and/or fibre properties, or having  
CC increased resistance to stresses or pests. Antisense, ribozyme or  
CC co-suppression molecules can be used to reduce the cellulose  
CC content of a transgenic plant, e.g. to improve digestibility or to  
CC alter carbon partitioning such that increased carbon is available  
CC for growth, rather than deposited as cellulose.  
CC  
XX  
XX Sequence 1065 AA:  
SQ  
Query Match 99.6%; Score 5655; DB 19; Length 1065;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MESEGTACKPMKNIVPOTCOICSDNVGKTVDDRFVACDSCFPVCPCEYERKDNQ 60  
DB 1 MESEGTACKPMKNIVPOTCOICSDNVGKTVDDRFVACDSCFPVCPCEYERKDNQ 60  
QY 61 SCPOCKTRKRLKSGPAIPGDDDEGLADGEVFEFYPCKEISERMGLHTRGKEEM 120  
DB 61 SCPOCKTRKRLKSGPAIPGDDDEGLADGEVFEFYPCKEISERMGLHTRGKEEM 120  
QY 121 GEPOYKEVSHNHLPLTSLRODTSGFSAASPERLSYSTIAGKRLPSSDVNOSPNNR 180  
DB 121 GEPOYKEVSHNHLPLTSLRODTSGFSAASPERLSYSTIAGKRLPSSDVNOSPNNR 180  
QY 181 IYDPVGLGNVAMKERDGMKOEKNTGVPSTOASERGVDDIDASTDILADALNDEA 240  
DB 181 IYDPVGLGNVAMKERDGMKOEKNTGVPSTOASERGVDDIDASTDILADALNDEA 240  
QY 241 KOPLSRKVSISSSRINPYRMVIMLRVILLCFLHYRITNPVNAFALMTVSVCEIMFAL 300  
DB 241 KOPLSRKVSISSSRINPYRMVIMLRVILLCFLHYRITNPVNAFALMTVSVCEIMFAL 300  
QY 301 SWILDOFPKWPVFNRTYIDRLALRYDRGEGPSQLAVIDFVSTVDPLEKPLVTANTYL 360  
DB 301 SWILDOFPKWPVFNRTYIDRLALRYDRGEGPSQLAVIDFVSTVDPLEKPLVTANTYL 360  
QY 361 SILAADVPAVDKVCYFDDGAAMLSPESLAETSEFAKWPVPCCKKYSIEBRAEWFAAK 420  
DB 361 SILAADVPAVDKVCYFDDGAAMLSPESLAETSEFAKWPVPCCKKYSIEBRAEWFAAK 420  
QY 421 IDYLDKQVTSFVKDRAKREVEEFKIRINMLVSKALCPBEGVWMDGTWPGNNTGD 480  
DB 421 IDYLDKQVTSFVKDRAKREVEEFKIRINMLVSKALCPBEGVWMDGTWPGNNTGD 480  
QY 481 HEGMTOVFLGONGGDAEGNELPRLVYVREKRPQOHKKKAGAMALYRVSAVLTNGSP 540  
DB 481 HEGMTOVFLGONGGDAEGNELPRLVYVREKRPQOHKKKAGAMALYRVSAVLTNGSP 540  
QY 541 IINLDCDHYINNSKALREAMCEFLMDPNLKOVCYVOPORFDGIDKNDYRANNTVEFDI 600  
DB 541 IINLDCDHYINNSKALREAMCEFLMDPNLKOVCYVOPORFDGIDKNDYRANNTVEFDI 600  
QY 601 NLRGLDGIQPYVGTGCVNRTALYGPPIKVKHKKPSLSKLCGSGRKKSKAKKSS 660  
DB 601 NLRGLDGIQPYVGTGCVNRTALYGPPIKVKHKKPSLSKLCGSGRKKSKAKKSS 660  
QY 661 DKKSGRHTDSTVPFNLDDIEBGVAGCFDDEKALLMSOMLEKRFQSAVFASTLME 720

DB 661 DKKSGRHTDSTVPFNLDDIEBGVAGCFDDEKALLMSOMLEKRFQSAVFASTLME 720  
QY 721 NGVPSAPPEENLKEAIIHIVISGVEDKSDMWMEIGMIYGSVEDILITGKMHARGMRST 780  
DB 721 NGVPSAPPEENLKEAIIHIVISGVEDKSDMWMEIGMIYGSVEDILITGKMHARGMRST 780  
QY 781 YCAPKLPAPFGSAPINLSDBLNQVLRWALGSVILFSRHCPYGYNGRLKLEPPAYVN 840  
DB 781 YCAPKLPAPFGSAPINLSDBLNQVLRWALGSVILFSRHCPYGYNGRLKLEPPAYVN 840  
QY 841 TTYIPTTSLPMLWCTLLAVCLFTNPIIPQISNIASIMFLSPLSIFATGILEMRSGV 900  
DB 841 TTYIPTTSLPMLWCTLLAVCLFTNPIIPQISNIASIMFLSPLSIFATGILEMRSGV 900  
QY 901 GIDEMRNNEQFWYIGVSAHLFAVFGIILKVLGIDTNFTVTSKASDEDDFAELYFKW 960  
DB 901 GIDEMRNNEQFWYIGVSAHLFAVFGIILKVLGIDTNFTVTSKASDEDDFAELYFKW 960  
QY 961 TLLIIPPTLLIYNLGVVAGSYAINSYGWGPFGKLFPAFWYIVHLYPELKGIMGR 1020  
DB 961 TLLIIPPTLLIYNLGVVAGSYAINSYGWGPFGKLFPAFWYIVHLYPELKGIMGR 1020  
QY 1021 QNRTPTIVYWSVLASISLSLWVRIDPFTSRVTGDIIECGINC 1065  
DB 1021 QNRTPTIVYWSVLASISLSLWVRIDPFTSRVTGDIIECGINC 1065  
RESULT 2  
AAC48733  
ID AAC48733 standard; Protein; 1065 AA.  
XX  
XX AAG48733;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 61573.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
OS termination sequence.  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 9905-0121825.  
XX  
XX 05-MAR-1999; 9905-0123180.  
XX  
XX 09-MAR-1999; 9905-0123548.  
XX  
XX 23-MAR-1999; 9905-0125788.  
XX  
XX 25-MAR-1999; 9905-0126264.  
XX  
XX 29-MAR-1999; 9905-0126785.  
XX  
XX 01-APR-1999; 9905-0127462.  
XX  
XX 06-APR-1999; 9905-0128234.  
XX  
XX 08-APR-1999; 9905-0128714.  
XX  
XX 16-APR-1999; 9905-0129845.  
XX  
XX 19-APR-1999; 9905-0130077.  
XX  
XX 21-APR-1999; 9905-0130449.  
XX  
XX 23-APR-1999; 9905-0130510.  
XX  
XX 28-APR-1999; 9905-0130891.  
XX  
XX 30-APR-1999; 9905-0131449.  
XX  
XX 30-APR-1999; 9905-0132048.  
XX  
XX 30-APR-1999; 9905-0132407.  
XX  
XX 04-MAY-1999; 9905-0132484.  
XX  
XX 05-MAY-1999; 9905-0132485.  
XX  
XX 06-MAY-1999; 9905-0132486.  
XX  
XX 07-MAY-1999; 9905-0132487.  
XX  
XX 11-MAY-1999; 9905-0132863.  
XX  
XX 11-MAY-1999; 9905-0134256.



PR 14-MAY-1999; 9905-0134218.  
PR 14-MAY-1999; 9905-0134219.  
PR 14-MAY-1999; 9905-0134221.  
PR 14-MAY-1999; 9905-0134370.  
PR 18-MAY-1999; 9905-01343768.  
PR 19-MAY-1999; 9905-0134941.  
PR 20-MAY-1999; 9905-0135124.  
PR 21-MAY-1999; 9905-0135353.  
PR 24-MAY-1999; 9905-0135629.  
PR 25-MAY-1999; 9905-0136021.  
PR 27-MAY-1999; 9905-0136382.  
PR 28-MAY-1999; 9905-0136782.  
PR 01-JUN-1999; 9905-0137222.  
PR 03-JUN-1999; 9905-0137528.  
PR 04-JUN-1999; 9905-0137502.  
PR 07-JUN-1999; 9905-0137724.  
PR 08-JUN-1999; 9905-0138094.  
PR 10-JUN-1999; 9905-0138540.  
PR 10-JUN-1999; 9905-0138847.  
PR 14-JUN-1999; 9905-0139119.  
PR 16-JUN-1999; 9905-0139452.  
PR 17-JUN-1999; 9905-0139453.  
PR 17-JUN-1999; 9905-0139492.  
PR 18-JUN-1999; 9905-0139454.  
PR 18-JUN-1999; 9905-0139455.  
PR 18-JUN-1999; 9905-0139456.  
PR 18-JUN-1999; 9905-0139457.  
PR 18-JUN-1999; 9905-0139458.  
PR 18-JUN-1999; 9905-0139459.  
PR 18-JUN-1999; 9905-0139460.  
PR 18-JUN-1999; 9905-0139461.  
PR 18-JUN-1999; 9905-0139462.  
PR 18-JUN-1999; 9905-0139463.  
PR 18-JUN-1999; 9905-0139750.  
PR 18-JUN-1999; 9905-0139763.  
PR 21-JUN-1999; 9905-0139817.  
PR 22-JUN-1999; 9905-0139899.  
PR 23-JUN-1999; 9905-0140353.  
PR 23-JUN-1999; 9905-0140354.  
PR 24-JUN-1999; 9905-0140685.  
PR 28-JUN-1999; 9905-0140823.  
PR 29-JUN-1999; 9905-0140991.  
PR 30-JUN-1999; 9905-0141287.  
PR 01-JUL-1999; 9905-0141842.  
PR 01-JUL-1999; 9905-0142154.  
PR 02-JUL-1999; 9905-0142055.  
PR 06-JUL-1999; 9905-0142380.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0142977.  
PR 13-JUL-1999; 9905-0143542.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 16-JUL-1999; 9905-0144086.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144352.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.

PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 02-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 09-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 01-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.  
PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158232.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159283.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.

PR 22-OCT-1999; 9905-0160989.  
 PR 25-OCT-1999; 9905-0161404.  
 PR 25-OCT-1999; 9905-0161405.  
 PR 25-OCT-1999; 9905-0161406.  
 PR 26-OCT-1999; 9905-0161359.  
 PR 26-OCT-1999; 9905-0161360.  
 PR 26-OCT-1999; 9905-0161361.  
 PR 28-OCT-1999; 9905-0161920.  
 PR 28-OCT-1999; 9905-0161922.  
 PR 28-OCT-1999; 9905-0161993.  
 PR 29-OCT-1999; 9905-0162142.

Query Match 99.6%; Score 5654; DB 21; Length 1065;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEESEGTACKPMKNIVPOTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNO 60  
 DB 1 MEESEGTACKPMKNIVPOTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNO 60  
 QY 61 SCPOCKTRYKRLKGSAPAIPEGKDEDEGLADGTYEENYPOKEKISEMHLGWHLTRKGGEEM 120  
 DB 61 SCPOCKTRYKRLKGSAPAIPEGKDEDEGLADGTYEENYPOKEKISEMHLGWHLTRKGGEEM 120  
 QY 121 GEPQYDKVESHNLPLRLTSRQDTSGEFSASPERLSVSTIAGKRLPYSSDVNOGSPNRR 180  
 DB 121 GEPQYDKVESHNLPLRLTSRQDTSGEFSASPERLSVSTIAGKRLPYSSDVNOGSPNRR 180  
 QY 181 IYDPVGLGNVAMKERDGMKMKOEKNTGPVSTQASERGVDDIDASTDILADALLNDEA 240  
 DB 181 IYDPVGLGNVAMKERDGMKMKOEKNTGPVSTQASERGVDDIDASTDILADALLNDEA 240  
 QY 241 RQPLSRKVSIPSSSRINPYRMVIMLRVLIICLFHYRITNPVPAFALMLVSVCEIEMFAL 300  
 DB 241 RQPLSRKVSIPSSSRINPYRMVIMLRVLIICLFHYRITNPVPAFALMLVSVCEIEMFAL 300  
 QY 301 SWILDOFPKWFPNVRETYLDRLALRYDREGSPQLAAVDIEVSTVDPLEKPLVTANTYL 360  
 DB 301 SWILDOFPKWFPNVRETYLDRLALRYDREGSPQLAAVDIEVSTVDPLEKPLVTANTYL 360  
 QY 361 SIILAVYYPVDKVCYVFDGGAAMLSESLAETSEFARKWPFCKKTSIEPRABEWYFAAK 420  
 DB 361 SIILAVYYPVDKVCYVFDGGAAMLSESLAETSEFARKWPFCKKTSIEPRABEWYFAAK 420  
 QY 421 IDYLKDKVOVSFYKDRRAKREYEERKIRINALVYSAKLCPEEGWVWODTTPPGNNTG 480  
 DB 421 IDYLKDKVOVSFYKDRRAKREYEERKIRINALVYSAKLCPEEGWVWODTTPPGNNTG 480  
 QY 481 HPGMIOVFLGONGGLDAEGNELPRLVYVSREKRPFQHHKKAGAMNALVVSAYLTNGPF 540  
 DB 481 HPGMIOVFLGONGGLDAEGNELPRLVYVSREKRPFQHHKKAGAMNALVVSAYLTNGPF 540  
 QY 541 IINLDDHYIINNSKALREAMCFIAMDNLKQVCYVQFPQRFQDIDKNDRAANNVYFPI 600  
 DB 541 IINLDDHYIINNSKALREAMCFIAMDNLKQVCYVQFPQRFQDIDKNDRAANNVYFPI 600  
 QY 601 NLNGLDIGIOPVYVGVCFNRTALGYEPPIKYKKKPSLKLKLGSGSKKSKAKKES 660  
 DB 601 NLNGLDIGIOPVYVGVCFNRTALGYEPPIKYKKKPSLKLKLGSGSKKSKAKKES 660  
 QY 661 DKKSGRHTDSTVPVNLDDIEGVEGAGFDEKALLMSOMLEKRRFGSAAVASTLME 720  
 DB 661 DKKSGRHTDSTVPVNLDDIEGVEGAGFDEKALLMSOMLEKRRFGSAAVASTLME 720  
 QY 721 NGGVPSATPENLTKRAIHVYISGVEYEDKSPMGMEIOWYISYVEDIILTGKMKARWRST 780  
 DB 721 NGGVPSATPENLTKRAIHVYISGVEYEDKSPMGMEIOWYISYVEDIILTGKMKARWRST 780  
 QY 781 YCMPIKLPARKGAPINILSDRLNOVLRMALGSVIELSRHCPIMYGYNGRLKFLERAAVY 840  
 DB 781 YCMPIKLPARKGAPINILSDRLNOVLRMALGSVIELSRHCPIMYGYNGRLKFLERAAVY 840  
 QY 841 TTIYPTISIPLLMYCTLLAVCLFTNOPIIPQISNIASIMWFLSLFISIFATGILEMRSGV 900

DB 841 TTIYPTISIPLLMYCTLLAVCLFTNOPIIPQISNIASIMWFLSLFISIFATGILEMRSGV 900  
 QY 901 GIDEMWRNQFWYIGVSAHLFAVPOGIIKVLAGITQNTVYSKADEGDFAEIYLFKM 960  
 DB 901 GIDEMWRNQFWYIGVSAHLFAVPOGIIKVLAGITQNTVYSKADEGDFAEIYLFKM 960  
 QY 961 TTIILIPPTLLIYNLGVAGVSAVAINSGVSGMPLGKLFPAFWIVHLHYPLKGLMR 1020  
 DB 961 TTIILIPPTLLIYNLGVAGVSAVAINSGVSGMPLGKLFPAFWIVHLHYPLKGLMR 1020  
 QY 1021 QNRTPTIVVWSVLASIFSLLMWRIDPFTSRVYTGPDIIIECGINC 1065  
 DB 1021 QNRTPTIVVWSVLASIFSLLMWRIDPFTSRVYTGPDIIIECGINC 1065

RESULT 3  
 ABB93356  
 ID ABB93356 standard; Protein; 1065 AA.  
 XX  
 AC ABB93356;  
 XX  
 DT 31-MAY-2002 (first entry)  
 DE Herbicidally active polypeptide SEQ ID NO 2567.  
 XX  
 KW Herbicidal; plant; agriculture; herbicide.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200210210-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PR 28-AUG-2001; 2001WO-EP09892.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Tietjen K, Weidner M.  
 XX  
 DR WPI; 2002-269010/31.  
 XX  
 PT Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -  
 XX  
 PS Claim 5; SEQ ID NO 2567; 261bp + Sequence Listing; English.  
 XX  
 CC The invention relates to identifying target proteins  
 CC (ABB90730-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.  
 CC  
 SQ Sequence 1065 AA;  
 XX  
 Query Match 99.6%; Score 5654; DB 23; Length 1065;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MEESEGTACKPMKNIVPOTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNO 60  
 DB 1 MEESEGTACKPMKNIVPOTCQICSDNVGKTVDGDRFVACDICSFPVCRPCYEYERKDGNO 60  
 QY 61 SCPOCKTRYKRLKGSAPAIPEGKDEDEGLADGTYEENYPOKEKISEMHLGWHLTRKGGEEM 120

Db 61 SCPOCKTRYKRLKGGSPAI PGDKEDGLADEGVTEFENYPOKEKISERMLGWHLTRKGGEEM 120  
QY 121 GEPQDKEVSHNHLRLTSRODTSGEFSAASPERLSVSTTGKRLPSPVNO SPNRR 180  
Db 121 GEPQDKEVSHNHLRLTSRODTSGEFSAASPERLSVSTTGKRLPSPVNO SPNRR 180  
QY 181 IVDPVGLGNVAMKEKERVDMKMKQEKNTGSPVSTQASERGVVIDIDASTDILADEALLNDEA 240  
Db 181 IVDPVGLGNVAMKEKERVDMKMKQEKNTGSPVSTQASERGVVIDIDASTDILADEALLNDEA 240  
QY 241 RQPLSRKVSIPSSRRINPYRMVIMLVILCLFLHRTINPVNAFALVSYICEIWEAL 300  
Db 241 RQPLSRKVSIPSSRRINPYRMVIMLVILCLFLHRTINPVNAFALVSYICEIWEAL 300  
QY 301 SWILDQFPKMPVFNNETYLDRLALRYDRGEPSSOLAADIENSTVDPLKEPPLVYANVL 360  
Db 301 SWILDQFPKMPVFNNETYLDRLALRYDRGEPSSOLAADIENSTVDPLKEPPLVYANVL 360  
QY 361 SILAVDYVDKVCYVFDGAAMLSFESELAETSEFARKWVPECKKYSIEPRAPEMYFAAK 420  
Db 361 SILAVDYVDKVCYVFDGAAMLSFESELAETSEFARKWVPECKKYSIEPRAPEMYFAAK 420  
QY 421 IDYLKDKVQTSFVKRRAMKREYEEFKIRINALVSKALCPREGEVVMODGTWPQNNITGD 480  
Db 421 IDYLKDKVQTSFVKRRAMKREYEEFKIRINALVSKALCPREGEVVMODGTWPQNNITGD 480  
QY 481 HPGMIOVFELGONGGLDAGNELPRLVYVYSREKRPFGOHKKGANNALVRSVAVLTNGFE 540  
Db 481 HPGMIOVFELGONGGLDAGNELPRLVYVYSREKRPFGOHKKGANNALVRSVAVLTNGFE 540  
QY 541 ILNLDCDHY INNSKALREAMCFLMDPNLKGQVCYVQFORFGIDKNDRYANRNVTFFDI 600  
Db 541 ILNLDCDHY INNSKALREAMCFLMDPNLKGQVCYVQFORFGIDKNDRYANRNVTFFDI 600  
QY 601 NLRGIDGQGVYVGTGCVFNRTALXGEPRITKVHKKRPSLSKLCGSSRKNSAKKES 660  
Db 601 NLRGIDGQGVYVGTGCVFNRTALXGEPRITKVHKKRPSLSKLCGSSRKNSAKKES 660  
QY 661 DKKSGRHTDSTVPFNLDIEEGEGAGFDEKALLSOMSELEKRFQGSAYFVASTLME 720  
Db 661 DKKSGRHTDSTVPFNLDIEEGEGAGFDEKALLSOMSELEKRFQGSAYFVASTLME 720  
QY 721 NGVPPSATPENLKEALIHVISCYGEDKSDMCMETIGVSYTEDILTGFKMHARGMSI 780  
Db 721 NGVPPSATPENLKEALIHVISCYGEDKSDMCMETIGVSYTEDILTGFKMHARGMSI 780  
QY 781 YCMPLPAFKGSAPINLSDRNLQVLRMALGSVEIIFSRHCRPTWYGNRKLKLERFAYVN 840  
Db 781 YCMPLPAFKGSAPINLSDRNLQVLRMALGSVEIIFSRHCRPTWYGNRKLKLERFAYVN 840  
QY 841 TTIVPTISIPILMCTLLAVCLFTNQFIIPQISNIASIWFSLFSTIFATGILEMRWGV 900  
Db 841 TTIVPTISIPILMCTLLAVCLFTNQFIIPQISNIASIWFSLFSTIFATGILEMRWGV 900  
QY 901 GIDEMWRNEQFWJIGVSANHLFAVQGIKVLGIDTFTVYSKASDEGDFAEALYLTKM 960  
Db 901 GIDEMWRNEQFWJIGVSANHLFAVQGIKVLGIDTFTVYSKASDEGDFAEALYLTKM 960  
QY 961 TTLILPPTLLIIVNLVGVAVSAINSGYSGMGLFEGKLPFAFVYIYHLFPFLGLMGR 1020  
Db 961 TTLILPPTLLIIVNLVGVAVSAINSGYSGMGLFEGKLPFAFVYIYHLFPFLGLMGR 1020  
QY 1021 QNRPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDILLECGINC 1065  
Db 1021 QNRPTIYVWMSVLLASIFSLMWRIDPFTSRVTPDILLECGINC 1065

RESULT 4  
ID AAG48734  
AC AAG48734;  
XX AAG48734;  
XX

DT 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 61574.  
DE Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS EP1033405-A2.  
PN  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 990S-0121825.  
XX 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 29-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130049.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 06-MAY-1999; 990S-0132487.  
PR 07-MAY-1999; 990S-0132863.  
PR 11-MAY-1999; 990S-0134256.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.  
PR 16-JUN-1999; 990S-0139453.  
PR 17-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139454.  
PR 18-JUN-1999; 990S-0139455.  
PR 18-JUN-1999; 990S-0139456.  
PR 18-JUN-1999; 990S-0139457.  
PR 18-JUN-1999; 990S-0139458.  
PR 18-JUN-1999; 990S-0139459.  
PR 18-JUN-1999; 990S-0139460.  
PR 18-JUN-1999; 990S-0139461.  
PR 18-JUN-1999; 990S-0139462.  
PR 18-JUN-1999; 990S-0139463.  
PR 18-JUN-1999; 990S-0139750.

PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139819.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142290.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161993.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 98.6%; Score 5597; DB 21; Length 1054;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1051; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 12 MKNIVPQTCICSDNKGTVGDRFVACDICSFPVCRPCYEXERKDNQSCPCQCKTRYR 71  
DB 1 MKNIVPQTCICSDNKGTVGDRFVACDICSFPVCRPCYEXERKDNQSCPCQCKTRYR 60

QY 72 LKGSFPAIPGDKDGEDGLADGTYEFNYPQEKISERLGNHLTRGKEENGEPQYDKREVS 131  
DB 61 LKGSFPAIPGDKDGEDGLADGTYEFNYPQEKISERLGNHLTRGKEENGEPQYDKREVS 120

QY 132 NMLPRLTSRODTSGEESASPRLSVSTIAGCKRLPYSSDVNOSPNNRRTVPVGLGNVA 191  
DB 121 NMLPRLTSRODTSGEESASPRLSVSTIAGCKRLPYSSDVNOSPNNRRTVPVGLGNVA 180

QY 192 MKERVDGWMKQEKNTGPVSTOASERGGVDIDASTDILADEALLNDEAROPLSRKVSIP 251  
DB 181 MKERVDGWMKQEKNTGPVSTOASERGGVDIDASTDILADEALLNDEAROPLSRKVSIP 240

QY 252 SSRINPYNVIMLRVILCLFLHYRTITNVPNAPFALMLVSVICETIFALSTIIDOPPKVF 311  
DB 252 SSRINPYNVIMLRVILCLFLHYRTITNVPNAPFALMLVSVICETIFALSTIIDOPPKVF 311

Db 241 SSRINPFRMVLRLVILCLFLHYRITNPVNAFMLVSYICEIWFALSWILDQFPKMF 300  
 Qy 312 PVRNRTYDLRLALRYDRGEPSSOLAADI FVSTVDPLKEPPLVNTANTVLSLADYDPDK 371  
 Db 301 PVRNRTYDLRLALRYDRGEPSSOLAADI FVSTVDPLKEPPLVNTANTVLSLADYDPDK 360  
 Qy 372 VSCYVFDGAAMLSFESIAETSEFARKVVPFCCKYSIETPRAPENYFAAKIDYLDKQVOTS 431  
 Db 361 VSCYVSDGAAMLSFESIAETSEFARKVVPFCCKYSIETPRAPENYFAAKIDYLDKQVOTS 420  
 Qy 432 FVKDRRAKREYEEFKIRINALVSKALCPREGVWMDGTWPNGNTGDHGMIOVFLGQ 491  
 Db 421 FVKDRRAKREYEEFKIRINALVSKALCPREGVWMDGTWPNGNTGDHGMIOVFLGQ 480  
 Qy 492 NGGLDAEGNELPRLYVYSREKRPFGQHHKAGANALVRSVAVLTNGFTINLDCDHTIN 551  
 Db 481 NGGLDAEGNELPRLYVYSREKRPFGQHHKAGANALVRSVAVLTNGFTINLDCDHTIN 540  
 Qy 552 NSKALREAMCFIAMPNLSKQCYVOFPORFGIDKNDRYRANRNTVFPIINLRGLDGIOP 611  
 Db 541 NSKALREAMCFIAMPNLSKQCYVOFPORFGIDKNDRYRANRNTVFPIINLRGLDGIOP 600  
 Qy 612 VYVGTGCVFNRTALYGEPIPIKVKHKPSLLSKLGGSRKKNSAKKESDKKSGRHTDS 671  
 Db 601 VYVGTGCVFNRTALYGEPIPIKVKHKPSLLSKLGGSRKKNSAKKESDKKSGRHTDS 660  
 Qy 672 TVPVPFNLDIEGVEGAFDEDEKALMSQMSLEKRFQGSANFVASTLMENGVPSAPPE 731  
 Db 661 TVPVPFNLDIEGVEGAFDEDEKALMSQMSLEKRFQGSANFVASTLMENGVPSAPPE 720  
 Qy 732 NLKLEAIHVISCYGEDSKDMGMEIGWITGVTEDILTFGKMHARQWRSIYCPKLPARKG 791  
 Db 721 NLKLEAIHVISCYGEDSKDMGMEIGWITGVTEDILTFGKMHARQWRSIYCPKLPARKG 780  
 Qy 792 SAPINLSDRNLQVLRMALGSVEILFSRHCPIMYGNRGLKFLERAPAVNTTIPYTSIPL 851  
 Db 781 SAPINLSDRNLQVLRMALGSVEILFSRHCPIMYGNRGLKFLERAPAVNTTIPYTSIPL 840  
 Qy 852 LMVCTILAVCLFTNOFTIPOISNASTIWELSLFISFTATGILEMWSGVGIDEMWRNQF 911  
 Db 841 LMVCTILAVCLFTNOFTIPOISNASTIWELSLFISFTATGILEMWSGVGIDEMWRNQF 900  
 Qy 912 WVIQGVSAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELIYFKWTLTLLIPTTLL 971  
 Db 901 WVIQGVSAHLFAVFOGILKVLAGIDNTFTVTSKASDEGDAELIYFKWTLTLLIPTTLL 960  
 Qy 972 IVNLVGVAVGVSYAINSQYQSMGPLFGKLFAPFVWIVHLVLPFLKGLMGRONRTPPIVVM 1031  
 Db 961 IVNLVGVAVGVSYAINSQYQSMGPLFGKLFAPFVWIVHLVLPFLKGLMGRONRTPPIVVM 1020  
 Qy 1032 SVLASTIFSLMVRIDPPTSRYVTPDILEGGINC 1065  
 Db 1021 SVLASTIFSLMVRIDPPTSRYVTPDILEGGINC 1054  
 RESULT 5  
 AAG48735  
 ID AAG48735 standard; Protein: 959 AA.  
 AAG48735;  
 18-OCT-2000 (first entry)  
 Arabidopsis thaliana protein fragment SEQ ID NO: 61575.  
 Protein identification; signal transduction pathway; metabolic pathway;  
 hybridisation assay; genetic mapping; gene expression control; promoter;  
 termination sequence.  
 Arabidopsis thaliana.  
 EP1033405-A2.

PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130044.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0132863.  
 PR 14-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139753.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 02-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 9905-0142390.  
PR 08-JUL-1999; 9905-0142803.  
PR 09-JUL-1999; 9905-0142920.  
PR 12-JUL-1999; 9905-0143277.  
PR 13-JUL-1999; 9905-0143342.  
PR 14-JUL-1999; 9905-0143624.  
PR 15-JUL-1999; 9905-0144005.  
PR 16-JUL-1999; 9905-0144085.  
PR 19-JUL-1999; 9905-0144325.  
PR 19-JUL-1999; 9905-0144331.  
PR 19-JUL-1999; 9905-0144332.  
PR 19-JUL-1999; 9905-0144333.  
PR 19-JUL-1999; 9905-0144334.  
PR 19-JUL-1999; 9905-0144335.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144632.  
PR 20-JUL-1999; 9905-0144884.  
PR 21-JUL-1999; 9905-0144814.  
PR 21-JUL-1999; 9905-0145086.  
PR 21-JUL-1999; 9905-0145088.  
PR 22-JUL-1999; 9905-0145085.  
PR 22-JUL-1999; 9905-0145087.  
PR 22-JUL-1999; 9905-0145089.  
PR 22-JUL-1999; 9905-0145192.  
PR 23-JUL-1999; 9905-0145145.  
PR 23-JUL-1999; 9905-0145218.  
PR 23-JUL-1999; 9905-0145224.  
PR 26-JUL-1999; 9905-0145276.  
PR 27-JUL-1999; 9905-0145913.  
PR 27-JUL-1999; 9905-0145918.  
PR 27-JUL-1999; 9905-0145919.  
PR 28-JUL-1999; 9905-0145951.  
PR 02-AUG-1999; 9905-0146386.  
PR 02-AUG-1999; 9905-0146388.  
PR 03-AUG-1999; 9905-0146389.  
PR 03-AUG-1999; 9905-0147038.  
PR 04-AUG-1999; 9905-0147204.  
PR 04-AUG-1999; 9905-0147302.  
PR 05-AUG-1999; 9905-0147192.  
PR 05-AUG-1999; 9905-0147260.  
PR 06-AUG-1999; 9905-0147303.  
PR 06-AUG-1999; 9905-0147416.  
PR 06-AUG-1999; 9905-0147493.  
PR 09-AUG-1999; 9905-0147935.  
PR 10-AUG-1999; 9905-0148171.  
PR 11-AUG-1999; 9905-0148319.  
PR 12-AUG-1999; 9905-0148341.  
PR 13-AUG-1999; 9905-0148565.  
PR 13-AUG-1999; 9905-0148684.  
PR 16-AUG-1999; 9905-0149368.  
PR 17-AUG-1999; 9905-0149175.  
PR 18-AUG-1999; 9905-0149426.  
PR 20-AUG-1999; 9905-0149722.  
PR 20-AUG-1999; 9905-0149723.  
PR 20-AUG-1999; 9905-0149929.  
PR 23-AUG-1999; 9905-0149902.  
PR 23-AUG-1999; 9905-0149930.  
PR 25-AUG-1999; 9905-0150566.  
PR 26-AUG-1999; 9905-0150884.  
PR 27-AUG-1999; 9905-0151065.  
PR 27-AUG-1999; 9905-0151066.  
PR 27-AUG-1999; 9905-0151080.  
PR 30-AUG-1999; 9905-0151303.  
PR 31-AUG-1999; 9905-0151438.  
PR 07-SEP-1999; 9905-0151930.  
PR 07-SEP-1999; 9905-0152363.  
PR 10-SEP-1999; 9905-0153070.  
PR 13-SEP-1999; 9905-0153758.  
PR 15-SEP-1999; 9905-0154018.  
PR 16-SEP-1999; 9905-0154039.  
PR 20-SEP-1999; 9905-0154779.  
PR 22-SEP-1999; 9905-0155139.

PR 23-SEP-1999; 9905-0155486.  
PR 24-SEP-1999; 9905-0155659.  
PR 28-SEP-1999; 9905-0156458.  
PR 29-SEP-1999; 9905-0156596.  
PR 04-OCT-1999; 9905-0157117.  
PR 05-OCT-1999; 9905-0157753.  
PR 06-OCT-1999; 9905-0157865.  
PR 07-OCT-1999; 9905-0158029.  
PR 08-OCT-1999; 9905-0158332.  
PR 12-OCT-1999; 9905-0158369.  
PR 13-OCT-1999; 9905-0159293.  
PR 13-OCT-1999; 9905-0159294.  
PR 13-OCT-1999; 9905-0159295.  
PR 14-OCT-1999; 9905-0159329.  
PR 14-OCT-1999; 9905-0159330.  
PR 14-OCT-1999; 9905-0159331.  
PR 14-OCT-1999; 9905-0159637.  
PR 14-OCT-1999; 9905-0159638.  
PR 18-OCT-1999; 9905-0159584.  
PR 21-OCT-1999; 9905-0160741.  
PR 21-OCT-1999; 9905-0160767.  
PR 21-OCT-1999; 9905-0160768.  
PR 21-OCT-1999; 9905-0160770.  
PR 21-OCT-1999; 9905-0160814.  
PR 21-OCT-1999; 9905-0160815.  
PR 22-OCT-1999; 9905-0160980.  
PR 22-OCT-1999; 9905-0160981.  
PR 22-OCT-1999; 9905-0160989.  
PR 25-OCT-1999; 9905-0161004.  
PR 25-OCT-1999; 9905-0161405.  
PR 25-OCT-1999; 9905-0161406.  
PR 26-OCT-1999; 9905-0161359.  
PR 26-OCT-1999; 9905-0161360.  
PR 26-OCT-1999; 9905-0161361.  
PR 28-OCT-1999; 9905-0161920.  
PR 28-OCT-1999; 9905-0161992.  
PR 28-OCT-1999; 9905-0161993.  
PR 29-OCT-1999; 9905-0162142.

Query Match 89.2%; Score 5066; DB 21; Length 959;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 956; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 107 MCGMHLTRKGEEMGEPOYDKVEYSHNHLRLTSLRODTSGEFSNAPSERLVSSTIAGKR 166  
DB 1 MCGMHLTRKGEEMGEPOYDKVEYSHNHLRLTSLRODTSGEFSNAPSERLVSSTIAGKR 60  
QY 167 LPPSSDVNOSPNNRITDVPGLGVNWKKEVDGKKMKOEKNTGVPVSOASBEGVDIDAS 226  
DB 61 LPPSSDVNOSPNNRITDVPGLGVNWKKEVDGKKMKOEKNTGVPVSOASBEGVDIDAS 120  
QY 227 TDLADEALLNDEAROPLSRKYSIPSSRIINPYRMVIMLRVILCLFLAHYRITNPVNAFA 286  
DB 121 TDLADEALLNDEAROPLSRKYSIPSSRIINPYRMVIMLRVILCLFLAHYRITNPVNAFA 180  
QY 287 LMLVSVICETWALSNILDOFPKFPVNBRETYLDRALRYDRBEGPSOLAANDIVSTVD 346  
DB 181 LMLVSVICETWALSNILDOFPKFPVNBRETYLDRALRYDRBEGPSOLAANDIVSTVD 240  
QY 347 PLKEPPLVANTVLSLAVDYVPDKSCVYFPDGAAMLSPESLAETSEFARKWVPCCKXY 406  
DB 241 PLKEPPLVANTVLSLAVDYVPDKSCVYFPDGAAMLSPESLAETSEFARKWVPCCKXY 300  
QY 407 SIEPRAPENYFAKIDYLDKQVTSFVKDRRAKREYEFKIRINALVSKALCPEEGVY 466  
DB 301 SIEPRAPENYFAKIDYLDKQVTSFVKDRRAKREYEFKIRINALVSKALCPEEGVY 360  
QY 467 MDGTPWPGNNTGDHDMGTOVFLGONGGDAEGNELPRLVYVSREKRPGEFHHKKAGAN 526  
DB 361 MDGTPWPGNNTGDHDMGTOVFLGONGGDAEGNELPRLVYVSREKRPGEFHHKKAGAN 420  
QY 527 ALYRVSALVTNGFFILNDCDHRINNASKALREAMCLAMPNIGKQYCYQFQRPBGIDK 586

D	b	422	ALVAVSALVINGPPIITLUDDCDTHINNSKALREAMCLMDPMLGQVCYVQRPQREJD	IDK	480
Q	y	587	NDRYANRNYVEFDINLRGLDGIQGPVYVGTGCVFNRTALVYGEPPIKYKHKRPSLSKLC		646
D	b	481	NDRYANRNTVEFDINLRGLDGIQGPVYVGTGCVFNRTALVYGEPPIKYKHKRPSLSKLC		540
Q	y	647	GGSKKKSKKAKKESDKKKSGRHNDSTVPVFNLDIIEBVEBAGPDDEKALLMSQMSLEKR		706
D	b	541	GGSKKKSKKAKKESDKKKSGRHNDSTVPVFNLDIIEBVEBAGPDDEKALLMSQMSLEKR		600
Q	y	707	FGOSAVEASTLMBENGVPSPSATENLKEAIVHISCGYEDKSPWGMETGVIYGSVTEDI		766
D	b	601	FGOSAVEASTLMBENGVPSPSATENLKEAIVHISCGYEDKSPWGMETGVIYGSVTEDI		660
Q	y	767	LITGKMHARGRSITCYMPKLPAFGSGAPINLSDNLQYLRRALDSVELLEFRHCPIMWGY		826
D	b	661	LITGKMHARGRSITCYMPKLPAFGSGAPINLSDNLQYLRRALDSVELLEFRHCPIMWGY		720
Q	y	827	NGRLKFLERFAYVNTTIYPIITSIPLMWCTLLAOCLEFNQFIPIQISNIASITWFLSFLS		886
D	b	721	NGRLKFLERFAYVNTTIYPIITSIPLMWCTLLAOCLEFNQFIPIQISNIASITWFLSFLS		780
Q	y	887	IFAGGILEMRSGVGDIMWRNEDPFWIIGYSAHLFAVFOGILKVIAGIDTINFYTSKAS		946
D	b	781	IFAGGILEMRSGVGDIMWRNEDPFWIIGYSAHLFAVFOGILKVIAGIDTINFYTSKAS		840
Q	y	947	DEDGDFAEIYLFKNTTLLIPIPTTLLIIVLVGVAVGSVAINSQVOSMGPLFGKLEFARVY		1008
D	b	841	DEDGDFAEIYLFKNTTLLIPIPTTLLIIVLVGVAVGSVAINSQVOSMGPLFGKLEFARVY		900
Q	y	1007	IVHLIYPLKGLMGQONRPTIIVVWVSVLASIFSLMLWRIDPFISRYTGPDILEEGINC		1065
		901	IVHLIYPLKGLMGQONRPTIIVVWVSVLASIFSLMLWRIDPFISRYTGPDILEEGINC		959

XX	RESULT 6
XX	AAV84108
ID	AAV84108 standard; Protein; 1077 AA.
XX	
AC	AAV84108;
XX	
DT	03-JUL-2000 (first entry)
DE	Amino acid sequence of a maize cellulose synthase.
XX	
KM	Maize; cellulose synthase; stalk quality; strand; silage; cellulose;
XX	transgenic plant; plant breeding marker.
OS	Zea mays.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 494
XX	/note= "encoded by MGS"
PN	MO200009706-A2.
PD	24-FEB-2000.
PF	16-AUG-1999; 99WO-US18760.
PR	17-AUG-1998; 98US-0096622.
PA	(PION-) PIONEER HI-BRED INT INC.
PI	Dhuga KS, Helentjaris TG, Bowen BA, Wang X;
DR	WPI; 2000-224343/19.
DR	N-PSDB; AAZ59494.
PT	New genes which encode maize cellulose synthase polypeptides in plants
PT	useful for modulating the expression of cellulose synthase in plants
PT	and to produce transgenic plants expressing the novel protein
IX	-

Claim 15; Page 95-97; 119pp; English.

The present sequence represents a maize cellulose synthase polypeptide. The cellulose synthase can be used for the improvement of stalk quality for improved stand or silage. It also provides an increased concentration of cellulose in the pecticarp, hardening the kernel and improving its handling ability. The sequences are used to produce transgenic plants and seeds expressing the cellulose synthase. The polynucleotide is used for modulating, preferably increasing, the level of the synthase in a plant cell. The plants are preferably monocots. The polynucleotide is also used as a probe or primer in the detection, quantitation or isolation of gene transcripts. The probes are useful in detecting deficiencies in the level of mRNA in screenings for desired transgenic plant, for detecting mutations in the gene, for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants of the gene, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides or for use as immunogens in the preparation and/or screening of antibodies. The proteins can be employed in assays for enzyme agonists or antagonists of enzyme function or for use of immunogens or antigens to obtain antibodies specifically immunoreactive with a protein.

Sequence 1077 AA;

Query Match	80.5%	Score 4572;	DB 21;	Length 1077;
Best Local Similarly	79.48%	Pred. No. 0;		
Matches 861; Conservative	81;	Mismatches 112;	Indels 30;	Gaps 10;

QY	4	EGETAG- KPMKINIVPOTCOICSDNNGKIVYDGRFACDISCFSPRCSPREYTEREKDGNOSC	62
Db	2	EGADADVKSQRREGGVCQICGDGVTTAEGVFAACVGCPRPCSPREYTEREKDGVTOAC	61
QY	63	POCKTRKRLKSGSPALIRGKDEQGLADECTVEFN-----POKISERMJGHLTRGK	116
Db	62	POCKTRKRLKHKSGSPALIRGEEDTDADS---DFNVLASGNEQOKLADPRKSRMNVG	115
QY	117	GEEBGEPOYDK-----EVSHNLPRLSHODTSGEBSASPERLSVSTIAGK	165
Db	119	SGGVGRPKYDSCGEIGLTKYDSEIRPGYIPYSTNQ--ISGEIPASPDHNMSPGTNIGK	177
QY	166	RLPESSDVYNSPRKRLVYDPVGLGANNAMKERVDGKMKDEKNTGPV---STOASBEGV-	221
Db	178	RAFPVY-VNHSPPSRSEFGSIGNAWAMKERDGMKMKDKCTIPTNTCTSIAPSEGRVC	236
QY	222	DIDASTDILADENLANDDEAROPLSRKVSIPSSRINPVMVIMLRLVILCLAHYRINPV	281
Db	237	DIDASTDYNNEDALLNDETRQPLSKRPVLPSSRINPVMVYLRILYVSTILNHRINPV	296
QY	282	PNAFALMLVSVICELINFALSWILDOFPKMFVYNRETYLDRILALRYDREGESPOLAAVDIF	341
Db	297	RNAVPLMLSVICELINFALSWILDOFPKMFPIRETYLDRILALRYDREGESPOLAAVDIF	356
QY	342	VSVAVDDLKBPPLVTANTVLSILAANDYVDVKVSCYVFDGGAAMLSPRESIAEFSFARAKMP	401
Db	357	VSVAVDPMKEBPPLVTANTVLSILAANDYVDVKVSCYSDGGAAMLFTDALAEFSFARAKMP	416
QY	402	FCCKYSIEPRAPEMYFAAKIDYLKDKVOTSEFYKDRRAKREYEEFKIRINMLVSKALCP	461
Db	417	FVKYKNIERAPREMYESQKIDYLKDKVHPSFYKDRRAKREYEEFKVAVNGLVAKQVP	476
QY	462	EEBWVWQDGTPMFGNNTGHPHMGIOVFLGONGGLDAEGNELPRLYUVYSREKRRPFGONHK	521
Db	477	EEBWIMQDGTPMFGNNTXHPGMIOVFLGSHSGGLDEGENELPRLYUVYSREKRRPFGONHK	536
QY	522	AGAMNALLVRSAYLITNGPFLIMLDDCHYINNSKALREAMCEAMPBNLQOYUQVOPORF	581
Db	537	AGAMNALLVRSAYLITNGOYMLNDDCHYINNSKALREAMCETLMBPNLGRSVCYQVOPORF	596
QY	582	DGIDKNDRYANRNTYFEDINLRGLDGIOPVYVGTGCVFNRTALXGEPRIKYVKKRPSL	641
Db	597	DGIDRNDRYANRNTYFEDINLRGLDGIOPVYVGTGCVFNRTALXGEPRI---POKSGE	654





QY 642 LSLKCGSRKKSKAKKESDKKSGRHTDSTVVFNLDDIEEGVAGFDEDEKALLMSQ 701  
DB 655 LSLKCGG-RKKASKSKKSGDKKSKQHVDSSVPFNLDEIEEGVAGFDEDEKALLMSQ 713  
QY 702 SLEKRRGQSAVFASTLMENGVPSPATPENLKEAIVHSCYEEDKSDMGEMIGWYGS 761  
DB 714 SLEKRRGQSAVFASTLMENGVPSPATPENLKEAIVHSCYEEDKTEWTEIGWYGS 773  
QY 762 VTEDDILTFGRMHARGRSIYCMPLKPAFGSAPINISDRNOYLKRNALGSVELLSRHP 821  
DB 774 VTEDDILTFGRMHARGRSIYCMPLKPAFGSAPINISDRNOYLKRNALGSVELLSRHP 833  
QY 822 IMVYNGRLKFLERFAYVNTIYPTISIPPLMYCTLLAVCLFNOPIIPDISNIASIMFL 881  
DB 834 LMVYNGRLKFLERFAYVNTIYPTISIPPLMYCTLLAVCLFNOPIIPDISNIASIMFL 893  
QY 882 SLEPLSFATGILEMRSGVIGIDEMWNEQFVWIGVSAHLFAVFOGILKYLACIDINFTY 941  
DB 894 SLEPLSFATGILEMRSGVIGIDEMWNEQFVWIGVSAHLFAVFOGILKYLACIDINFTY 953  
QY 942 TSKASDEDDGFAELTYFKMTTLIPPTLLIIVNLGVAGVSAINSYOSMGPLGKLF 1001  
DB 954 TSKASDEDDGFAELTYFKMTTLIPPTLLIIVNLGVAGVSAINSYOSMGPLGKLF 1013  
QY 1002 FAFWIVHLXPLKGLMGRNRTPTIYVWVSYLASFSLMWRIDPFTSRVGPDIIEC 1061  
DB 1014 FAFWIVHLXPLKGLMGRNRTPTIYVWVSYLASFSLMWRIDPFTSRVGPDIIEC 1073  
QY 1062 GINC 1065  
DB 1074 GINC 1077

## RESULT 8

AAU10496  
ID AAU10496 standard; Protein; 1079 AA.

XX AAU10496:

DT 14-FEB-2002 (first entry)

DE Corn cellulose synthase Cqrae19/cesa-9.

XX Corn; cellulose synthase; Cdpgs45; cesa-3; Cqrae19; cesa-9;

KW stalk quality; improved stand; silage; pericarp; kernel hardening;

XX handling ability; transgenic plant; immunogen.

OS Zea mays.

XX WO200179516-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11951.

XX 14-APR-2000; 2000US-0550483.

XX (PION-) PIONEER HI-BRED INT INC.

XX Dhuga KS, Helemtjaris TG;

XX WPI: 2002-041338/05.

XX N-PSDB: AAS16458.

XX New cellulose synthase polypeptides and polynucleotides, useful in

XX improving stalk quality or silage, and in increasing concentration of

XX cellulose in the pericarp, hardening the kernel for improved handling

XX ability

XX Claim 3; Page 85-87; 88pp; English.

XX The invention relates to isolated nucleic acids encoding two cellulose

XX synthase proteins from corn, Cdpgs45 (cesa-3) and Cqrae19 (cesa-9).

CC Also disclosed are a recombinant expression cassette comprising the  
CC polynucleotide (operably linked to a promoter) a host cell comprising  
CC the recombinant expression cassette and a transgenic plant comprising  
CC the recombinant expression cassette. The nucleic acid is useful in  
CC the improvement of stalk quality for improved stand or silage, and in  
CC the improved concentration of cellulose in the pericarp, hardening  
CC the kernel, and thus improving its handling ability. The nucleic acids  
CC may also be used as probes or amplification primers in the detection,  
CC quantification or isolation of gene transcripts, as probes in detecting  
CC deficiencies in the level of mRNA, for detecting gene mutations or  
CC allelic variants, for monitoring up regulation of expression or changes  
CC in enzyme activity in screening assays, for site directed mutagenesis,  
CC in sense or antisense suppression of one or more genes in a host  
CC cell, tissue or plant. The polypeptides may be used in assays for  
CC enzyme agonists or antagonists, as immunogen or antigen to  
CC obtain antibodies specifically immunoreactive with the protein.  
CC The present sequence represents a corn cellulose synthase of the  
CC invention.

SQ Sequence 1079 AA;

Query Match 80.4%; Score 4564; DB 23; Length 1079;

Best Local Similarity 79.1%; Pred. No. 0;

Matches 858; Conservative 86; Mismatches 111; Indels 30; Gaps 11;

4 EGFTAG-KPMKNIVPTCCICSDNVCVTVDGDFVACDICSFVCRPCYEERKDNOSC 62

2 EGDADGVKSGRRGGVQICGQVGTAGDFVACDVCGFVCRPCYEERKDNOSC 61

63 PCKTRYKRLKGSFPAIGDKDEGLADEGTEFENYF-----QEKISERMLQMHLLRGR 116

62 PCKKNYKRLKGSFPAIRGEEDDTADDA--DENYFASGNDQKRIADMRSMNAG 120

117 GEEKGEPOYDK-----EVSNNHLPRLTSKODTSGEFAASPERLSVSTTAGK 165

121 SGDVGRPKYDSGEIGLTKYDSGEIPKRYIPSVTNSQ--ISGEIGASPSDHMMSPNTNIGR 179

166 RLPLYS--SDVNSQSPNRRLVDPVGLGVNAAKRVGMKKOKKNGPV---STOASERGV 221

180 RAFPYNNHSSNPSREPSGV--GVNAKKRVGMKKOKKGTIPMNGTINPSSEGRV 237

222 -DIDASTDLADEALLNDEARQPLSRKVSIPSSRINPYRVNMLRLVILCLFLAYRTNP 280

238 GDIDASTDYMEALLNDEARQPLSRKVSIPSSRINPYRVNMLRLVILCLFLAYRTNP 297

281 VPAFALMLVSVICELFALSWILDQPKKFPVNNRETYLDRLALRDRSEPSQLAVDI 340

298 VRNAYPLMLSVICELFALSWILDQPKKFPVNNRETYLDRLALRDRSEPSQLAVDI 357

341 FVSTVPLKEPPLVTANTVLSILAVDPYVNVKSCYVDDGAALSPESLAESEFARKW 400

358 FVSTVDMKEPPLVTANTVLSILAVDPYVNVKSCYVDDGAALSPESLAESEFARKW 417

401 PCKKYSIEPRAPDEWYFAAKIDYLDKQVQTSFVKDRAMKREYEERFIRNALVSALKC 460

418 PFVKKYNIERAPDEWYFSQIDYLDKQVHPSFVKDRAMKREYEERFIRNGLVAAQKV 477

461 PEEGWYMODGTPWPGNNTGHPGMIOVFLQNGGLDAEGNELPRLVYVSEKRRPGOHK 520

478 PEEGWYMODGTPWPGNNTGHPGMIOVFLQNGGLDTEGNEPRLVYVSEKRRPGOHK 537

521 KAGAMNVLVSVAVLNGPFLMLDCDHYNNKSKALREACFLMDPRLGQVGYVOPOR 580

538 KAGAMNVLVSVAVLNGPFLMLDCDHYNNKSKALREACFLMDPRLGQVGYVOPOR 597

581 FDGIDKNDRYANNTVEFFDINLGLDIOGPVYVGTGCVFNRTALGYEPPIKVKHKKS 640

598 FDGIDKNDRYANNTVEFFDINLGLDIOGPVYVGTGCVFNRTALGYEPPI--KQKKG 655

641 LSLKCGSRKKSKAKKESDKKSGRHTDSTVVFNLDDIEEGVAGFDEDEKALLMSQ 700

656 LSLKCGG-RKKASKSKKSGDKKSKQHVDSSVPFNLDEIEEGVAGFDEDEKALLMSQ 714



[illegible]

Query Match	77.8%	Score 4418	DB 21	Length 1043
Best Local Similarity	79.3%	Pred. No. 0		
Matches 833	Conservative 78	Mismatches 109	Indels 30	Gaps 10
QY	4	EEETG-G-KPMKIVVQTCIGSDNKGKTYVDDGRFPAACDICSPPVRCPEYERKAGNSGC	62	
DB	2	EEDAGGVSVGRGGGGOVCIGDGGTGAEGVAFACDVCSPVRCPEYERKAGTQAC	61	
QY	63	POCKTKRYRLKSPAIIPGDKDEGLADEGTEFENY-----POKEKISERMLGWLTRK	116	
DB	62	POCKTKRYRKHKSPAIRGEGDDTDADS---DFNTLAGSNEQOKIADRMKSMRYNNGG	118	
QY	117	GEEMGEPOYDK-----EVSHNHLRLTSSKODTSGERSAASPERLSVSYTAGK	165	
DB	119	SGDVERPKRDKDSEIGLTKRKDSEIRGRYIPSTNSQ-LSGELPGASPDHMMSPIGNIGK	177	
QY	166	RLPYSSDVNQSNNRIIVDPVGLGNVAMKERVDMKMKQEKNTGPV---STQASERGV-	221	
DB	178	RAFPFY-VNHSNPSPREFSGSIGNAVAMKERVGMKMKQDKGTIPMTNGTSIAPSEKGVG	236	
QY	222	DIDASTDILADALLDEAROPLSKRSVPSRRINPBYRVMLRLVYLCTFLHYRTNPV	281	
DB	237	DIDASTDINMEDALLNDETROPLSKRVPLPSSRINPBYRVMLRLVLSIFHYRTNPV	296	
QY	282	PNAFPLMTLVSVCEIMFALSWILDOFPKPFVPVNRRETYIDRLRLRYDREGEPSQLAVIDF	341	
DB	297	RAAYPLMWLTSVCEIMFALSWILDOFPKPFVNRRETYIDRLRLRYDREGEPSQLAVIDF	356	
QY	342	VSTVDPLKEPPLVTANFVLSTIAVDY PVDKVCYVFDDGAMLSFEBSLAETISEPARKVP	401	
DB	357	VSTVPMKEPPLVTANTVLSIAVDY PVDKVCYVSDDAAMLTFDALAETISEFARKVP	416	
QY	402	FOKKKSIIEPRAEEMVFAKIDYLKQVQSFYKDKBRAMKEVEERKIRINALVSAALCP	461	
DB	417	FAKKKNIIEPRAEEMVFSQIDYLKQVHSPFYKDKBRAMKEVEEVRVNGVLAQAQKP	476	
QY	462	EEGVMVODCTPMPGNNTGDHPGMIOVFLQNGCLDAEGNELRLVYVSEKRPGEONHK	521	
DB	477	EEGVMVODCTPMPGNNTGDHPGMIOVFLQNGCLDTEGNEELRLVYVSEKRPGEONHK	536	
QY	522	AGAMNALVVSALVTNGEPFILNDCDHYTINNSKALREAMCFLMDPILQKQCVOPORF	581	
DB	537	AGAMNALVVSALVTNGQYMLNDCDHYTINNSKALREAMCFLMDPILQKQCVOPORF	596	
QY	582	DOIDKNDKRRANNYVFPDNLNLGDLGIGQPVYVVGCGVNNRNLGVEPPIVKKKKPSL	641	
DB	597	DOIDKNDKRRANNYVFFDNLNLGDLGIGQPVYVVGCGVNNRNLGVEPPIVKKKKPSL	654	
QY	642	LSKLCGSGRRKKSKAKKESDKKSRHTQSTVPVFLDIEEGVAGAGDEDEKALMSQM	701	
DB	655	LSKLCGSG-RKRAKSKKSGDKKKSKQKHDSVPVFLNEDIEEGVAGAGDEDEKALMSQM	713	
QY	702	SLERKFGOSAVFVASTLMEGCVPPSPATPENLLKRAIHVISCYGEDKSDMGMEIGMTGS	761	

DB 714 SLEKRFQSAFVASTIMEYGVGVSQATSEBSLKEAIIHVISCYEDKTEMGTEIGNIYGS 773  
QY 762 VTEEDILTGFKMARGRSITICMPKLPKFAFGSAPINLSDRLNOVLRWALGSVELTFRHCP 821  
DB 774 VTEEDILTGFKMARGRSITICMPKLPKFAFGSAPINLSDRLNOVLRWALGSVELTFRHCP 833  
QY 822 IWYNGRLKFLERFAYVNTTITPITISPLMYCTLLAVCLFTNPTIPQISNIASIMPL 881  
DB 834 LHWGYGRKFLERFAYVNTTITPITISPLMYCTLLAVCLFTNPTIPQISNIASIMPL 893  
QY 882 SLELSTFATGILEMRSSGVIDEMWNEQFVWIGSAHLFAVFGSLVLAAGIDITNFTV 941  
DB 894 SLELSTFATGILEMRSSGVIDEMWNEQFVWIGSAHLFAVFGSLVLAAGIDITNFTV 953  
QY 942 TSKASDEDDFAELLYFKMTTLLIPTTLLIVNLVGVVAGVSAINSQVSMGPLEGKLF 1001  
DB 954 TSKASDEDDFAELLYFKMTTLLIPTTLLIVNLVGVVAGVSAINSQVSMGPLEGKLF 1013  
QY 1002 FAFWIVHLVLPFLKGLMGQRNPTPIVVW 1031  
DB 1014 FAFWIVHLVLPFLKGLMGQRNPTPIVVW 1043

RESULT 11  
AA184109  
ID AA184109 standard; Protein; 1075 AA.  
XX  
AC AA184109;  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE Amino acid sequence of a maize cellulose synthase.  
XX  
KM Maize: cellulose synthase; stalk quality; stand; silage; cellulose;  
KM transgenic plant; plant breeding marker.  
XX  
OS Zea mays.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 245  
XX /note= "encoded by NAT"  
XX  
PN WO200009706-A2.  
PD 24-FEB-2000.  
XX  
PF 16-AUG-1999; 99WO-US18760.  
PR 17-AUG-1998; 98US-0096822.  
XX  
PA (PION-) PIONEER HI-BRED INT INC.  
XX  
PI Dhuga KS, Helentjaris TG, Bowen BA, Wang X;  
XX  
DR WPI, 2000-224343/19.  
XX N-PSDB: AA294947.  
XX  
PT New genes which encode maize cellulose synthase polypeptides in plants  
PT useful for modulating the expression of cellulose synthase in plants  
PT and to produce transgenic plants expressing the novel protein  
XX  
XX Claim 15; Page 102-105; 119pp; English.  
XX  
XX The present sequence represents a maize cellulose synthase polypeptide.  
CC The cellulose synthase can be used for the improvement of stalk quality  
CC for improved stand or silage. It also provides an increased concentration  
CC of cellulose in the pericarp, hardening the kernel and improving its  
CC handling ability. The sequences are used to produce transgenic plants  
CC and seeds expressing the cellulose synthase. The polynucleotide is  
CC used for modulating, preferably increasing, the level of the synthase  
CC in a plant cell. The plants are preferably monocots. The polynucleotide  
CC is also used as a probe or primer in the detection quantitation or  
CC isolation of gene transcripts. The probes are useful in detecting

CC deficiencies in the level of mRNA in screenings for desired transgenic  
CC plant, for detecting mutations in the gene, for monitoring upregulation  
CC of expression or changes in enzyme activity in screening assays of  
CC compounds, for detection of any number of allelic variants of the gene,  
CC or for use as molecular markers in plant breeding programs. The  
CC isolated nucleic acids of the present invention can also be used for  
CC recombinant expression of their encoded polypeptides or for use as  
CC immunogens in the preparation and/or screening of antibodies. The  
CC proteins can be employed in assays for enzyme agonists or antagonists  
CC of enzyme function or for use of immunogens or antigens to obtain  
CC antibodies specifically immunoreactive with a protein.  
XX  
XX Sequence 1075 AA:  
QY Query Match 71.0%; Score 4028.5; DB 21; Length 1075;  
Best Local Similarity 69.6%; Pred. No. 0;  
Matches 752; Conservative 130; Mismatches 153; Indels 45; Gaps 12;  
DB 4 EGETAG--KPMKNIVPQTQICSDNYGKTVDGDRVACDICSFPVPCPCYERKDGNS 61  
23 DGDVPGSAKPTKSANQVQCIGDSVGSATGDVFAVACNECAFPCPCYERKDGNS 82  
QY 62 CPQCKTRYKRLGSPAIPGDKDEGLADGTYEPNTPQCKTISENMLGRGGEENG 121  
DB 83 CPQCKTRYKRLGSPAIPGDKDEGLADGTYEPNTPQCKTISENMLGRGGEENG 128  
QY 122 EPQYDKES-----HNHLPLRUTSRQDTSGEFSASPERLSVSTIAGKRLPYSSDVN 174  
DB 129 LOGDDADLSSASRHEPHHRIPRLTSGQQLSGELPDASPRHSI-----KSPISSYVD 180  
QY 175 QS--PNKRIVDP-----VGLGNVAKKERVNDGKMKOEKNTGPVSTQAASERGVDIDAS 226  
DB 181 PSVPVPRVIVDSKDLNSYGLNSVDWKKEVESWRVQDKNMQVINYKYPEARAG--DME-G 238  
QY 227 TDILADEALLNDENARPLSRKYSIPSSRINPRMYIMLRVILCLFHLIRINPPNMA 286  
DB 239 TGSNGEXMQVADVADRLPSRIVIPISSNOLNLRVILILLLCFEFOIRVSHPRDAG 298  
QY 287 LMLVSVICELPFWALSWILDQFPKMPFNRETLDRLALRYDREGESQILAADIVSTVD 346  
DB 299 LMLVSVICELPFWALSWILDQFPKMPFNRETLDRLALRYDREGESQILAADIVSTVD 358  
QY 347 PLKEPPLVANTVLSILAVDYVDVKVSCYFEDGAAMLSEBSLAETSEFAKRWVPCKKY 406  
DB 359 PLKEPPLVANTVLSILAVDYVDVKVSCYFEDGAAMLSEBSLAETSEFAKRWVPCKKH 418  
QY 407 STEPRAPEMYFAKIDYLDKQVOTSPVQRRAMKREYEFKIRIINALVAKLCKPEEGV 466  
DB 419 NIEPRAPERYFAKIDYLDKQVOTSPVQRRAMKREYEFKIRIINALVAKLCKPEEGV 478  
QY 467 MODGTWPNGNNTGDHFGMIQVFLGONGGLDAGNELPRLVYVSRKRPFGQHHKKRAGAN 526  
DB 479 MADGTAMPNPNRDRHGMIOVFLGSHSGGLDGTGNELPRLVYVSRKRPFGQHHKKRAGAN 538  
QY 527 ALVRVSAVLTNGPFLINDCDHYINNSKALREAPMGFLDPNIGKOVCTYQFORBDGIDK 586  
DB 539 ALIRVSAVLTNGAYLLNDVCDHYFNSKALREAPMGFLDPNIGKOVCTYQFORBDGIDL 598  
QY 587 NDRVYANRNVFEDINLRGIDGIGQRPVYTGCVFNRTALYGEPRKYKHKPSLSKLK 646  
DB 599 HDRVYANRNVFEDINLRGIDGIGQRPVYTGCVFNRTALYGEPRKYKHKPSLSKLK 658  
QY 647 GGSRRKSKAKKESDKSGRSTDSVPVFNLDIEEGVAGFDERKALMSQMSLEK 706  
DB 659 GGRKSKKNSYNDQSRIMKR--TESAPIFNMEDIEGIE--GYDERSVLMSQKLEK 715  
QY 707 PQGSAAVASTIMENGVPSPATPENLKEAIIHVISCYEDKTEMGTEIGNIYGSYTDI 766  
DB 716 PQGSAAVASTIMENGVPSPATPENLKEAIIHVISCYEDKTEMGTEIGNIYGSYTDI 775  
QY 767 LTFGFKHARGWOSITICMPKLPKFAFGSAPINLSDRLNOVLRWALGSVELTFRHCP 826  
DB 776 LTFGFKHARGWOSITICMPKLPKFAFGSAPINLSDRLNOVLRWALGSVELTFRHCP 835

QY	827	NGRLKLEFRFAYVMTTPTPTISIPILAMCTLLVACLFNQTIOISINASIAPLSELS	886		
Db	836	NGRLKLEFRLAYINTTPTISIPILACVLPALCLLNKFIPEISNYAGMEFTLLPAS	895		
QY	887	IFATGILEMRMSGVIGIDEMWRNEQFMWYGVGSAAHLFAVFOGILKVLACIDINFTVTSKAS	946		
Db	896	IFATGILELRMSGVIGIDEMWRNEQFMWYIGGSAHLFAVFOGLKVLACIDINFTVTSKAS	955		
QY	947	DEDGDFAEFLYLFKWTLLIPPTLLIYNLVGVAVGSVAINSQSGWSGPLEFGKLEFAFMV	1006		
Db	956	DEDGDFAEFLYLFKWTSLIIPPTLYLVINLVGMVAGISVAINSQSGWSGPLEFGKLEFSIMV	1015		
QY	1007	IVHLYPFLKGLMGNGRNPRTTIVWWSVLLASIFSLTWRIIDPFPSRYVGPDL-ECGINC	1065		
Db	1016	ILHLYPFLKGLMGNGRNPRTTIVWWSVLLASIFSLTWKIDPFISPTQKAAALGCGVNC	1079		
RESULT 12					
XX	ID	AAV84115	AAV84115		
XX	AC	AAV84115	standard; Protein; 1075 AA.		
XX	DF	03-JUL-2000	(first entry)		
XX	DE	Amino acid sequence of a maize cellulose synthase.			
XX	KW	Maize; cellulose synthase; stalk quality; stand; silage; cellulose;			
XX	RW	transgenic plant; plant breeding marker.			
XX	OS	Zea mays.			
FX	FE	Key	Location/Qualifiers		
FT	Misc-difference	245	/note= "encoded by NAT"		
XX	PN	WO200009706-A2.			
XX	PD	24-FEB-2000.			
XX	PF	16-AUG-1999;	99WO-US18760.		
XX	PR	17-AUG-1998;	98US-0096822.		
PA	(PION-) PIONEER HI-BRED INT INC.				
XX	PI	Dhugga KS, Helentjaris TG, Bowen BA, Wang X;			
XX	DR	WPI; 2000-224343/19.			
XX	DR	N-PSDB; AA299515.			
XX	PT	New genes which encode maize cellulose synthase polypeptides in plants			
XX	PT	useful for modulating the expression of cellulose synthase in plants			
XX	PT	and to produce transgenic plants expressing the novel protein			
PS	Claim 15; Page 149-152; 119pp; English.				
CC	The present sequence represents a maize cellulose synthase polypeptide.				
CC	The cellulose synthase can be used for the improvement of stalk quality				
CC	for improved stand or silage. It also provides an increased concentration				
CC	of cellulose in the pericarp, hardening the kernel and improving its				
CC	handling ability. The sequences are used to produce transgenic plants				
CC	and seeds expressing the cellulose synthase. The polynucleotide is				
CC	used for modulating, preferably increasing, the level of the synthase				
CC	in a plant cell. The plants are preferably monocots. The polynucleotide				
CC	is also used as a probe or primer in the detection quantitation or				
CC	isolation of gene transcripts. The probes are useful in detecting				
CC	deficiencies in the level of mRNA in screenings for desired transgenic				
CC	plant, for detecting mutations in the gene, for monitoring upregulation				
CC	of expression or changes in enzyme activity in screening assays of				
CC	compounds, for detection of any number of allelic variants of the gene,				
CC	or for use as molecular markers in plant breeding programs. The				

QY	887	FEANGILEMRSSGVGIDEMWNNEDFPWYTGVSAAHFAFQGLIKVLAVGIDTNPVTSKAS	946
DB	896	IFANGILELRNNSGVGIEDMWNEDQFWYGGTSAHLFAVFGSLKVLAVGIDTNPVTSKAS	955
QY	947	DEDDGFAELVYEFKTTLLIPTTLLIYLVGVVAVGSVAINSGYOSWGPLEGKLEFAFW	1006
DB	956	DEDDGFAELVYEFKTTLLIPTTLLIYLVGVVAVGSVAINSGYOSWGPLEGKLEFSIMV	1015
QY	1007	IVHLVYPLFKGLMGQRNPRTIVVYVWSVLLAIFSLMLWRIDPFNSRYTGPIIL-ECGINC	10655
DB	1016	ILHLVPLFKGLMGQRNPRTIVVYVWSVLLAIFSLMLWRIDPFISIPQKAALQCCGVNC	1075
RESULT 13			
ID	AAV84120		
XX	AAV84120	standard; Protein: 1075 AA.	
AC	AAV84120;		
XX	03-JUL-2000	(first entry)	
DE		Amino acid sequence of a maize cellulose synthase.	
XX			
KM		Maize; cellulose synthase; stalk quality; stand; silage; cellulose;	
KW		transgenic plant; plant breeding marker.	
XX			
OS		Zea mays.	
XX			
FT	Key	location/qualifiers	
FT	Misc-difference 245	/note= "encoded by NAT"	
PN		WO200009706-A2.	
XX			
PD	24-FEB-2000.		
XX			
PE	16-AUG-1999;	99WO-US18760.	
XX			
PR	17-AUG-1998;	98US-0096822.	
XX			
PA	(PION-) PIONEER HI-BRED INT INC.		
XX			
PI	Dhugga KS, Helentjaris TG, Bowen BA, Wang X;		
XX			
DR	WPI: 2000-224343/19.		
XX			
DR	N-PSDB: AA299530.		
XX			
PS			
XX	Claim 15; Page 188-191; 119pp; English.		
CC			
CC	The present sequence represents a maize cellulose synthase polypeptide.		
CC	The cellulose synthase can be used for the improvement of stalk quality		
CC	for improved stand or silage. It also provides an increased concentration		
CC	of cellulose in the pericarp, hardening the kernel and improving its		
CC	handling ability. The sequences are used to produce transgenic plants		
CC	and seeds expressing the cellulose synthase. The polynucleotide is		
CC	used for modulating, preferably increasing, the level of the synthase		
CC	in a plant cell. The plants are preferably monocots. The polynucleotide		
CC	is also used as a probe or primer in the detection quantitation or		
CC	isolation of gene transcripts. The probes are useful in detecting		
CC	deficiencies in the level of mRNA in screenings for desired transgenic		
CC	plant, for detecting mutations in the gene, for monitoring upregulation		
CC	of expression or changes in enzyme activity in screening assays of		
CC	compounds, for detection of any number of allelic variants of the gene,		
CC	or for use as molecular markers in plant breeding programs. The		
CC	isolated nucleic acids of the present invention can also be used for		
CC	recombinant expression of their encoded polypeptides or for use as		
CC	immunogens in the preparation and/or screening of antibodies. The		
CC	proteins can be employed in assays for enzyme agonists or antagonists		
CC	of enzyme function or for use of immunogens or antigens to obtain		
CC			

CC	antibodies specifically immunoreactive with a protein.
XX	Sequence 1075 AA;
Query Match	71.0%; Score 4028.5; DB 21; Length 1075;
Best Local Similarity	69.6%; Pred. No. 0;
Matches	752; Conservative 130; Mismatches 153; Indels 45; Gaps 12;
QY	4 EGETAG--KPKAKNIYPCOTICSDNKVTGVGDRLVACDDISFPYCRRCYERYKDGMS 61
DB	23 DGDVPGSAKPKRSANGQVCIGSVGASGADVFAVACNECAPFCRCYCEYERKEGMC 82
QY	62 CPCKTRKRLKGSFAIPGDDEGLADEGVTEFNYPQEKERISERMLGMHLTRGKEEMG 121
DB	83 CPCKTRKRLKGSFRAVHGDEDEEV--DDLNEFNRYKGS-----GKGPMMO 128
QY	122 EPQYDKEYS-----INHLPRLTQRQTSSEFSAASPERLSVSTAGKRLPYSSDVN 174
DB	129 LOGDDADDSARHEPHRIPLRTLSGOQISGEIIPASDRAHSI-----RSPSYSDV 180
QY	175 QS--PNRRIVP-----VGLGNVAMKRRYGVGMKKQKNPGVSTOASERGGVDIDAS 226
DB	181 PSVPYPAIVDPKSDLNSTGLNSVDKRRVSWRKQDKMMQVNTNKIPKARG--DME--G 238
QY	227 TDILADEALLDEAKROPLSRKRVSPSSRINERYRMYIMLRVILLCFLHRYITNPVNAFA 286
DB	229 TGSNGEXMQWDDARLPISRIVPTISSNOANTLYRVLITRLILLCFFQYRSHVPYDAAG 298
QY	287 LMLVSVICELFALSMILDDPPKPFVPRARETYDLRLARYREGSPOLAVIDFVSTVD 346
DB	299 LMLVSVICELFALSMILDDPPKWPPIRRETYDLRLARYREGSPOLAVIDFVSTVD 358
QY	347 PLKEPPLTANTVLSILAADVPRVDKVSQCYVDDGAAMLSFESLAESEFARKWPFCKKY 406
DB	359 PLKEPPLTANTVLSILSVDPYVDKVSQCYVDDGSAMLTFSLSPTAFARKWPFCKKH 418
QY	407 SIERRAPWYTAAKIDYLLKDKVQVTSFVKORAMKREYEEKIRINALYSKALCPSEGVY 466
DB	419 NIEPRAPETFYAOKIDYLLKDKIOPSFYERARKREYEEFVRINALVAKQKYEDEMT 478
QY	467 MODGPMPNGNNTGHPGMIOYFELONGGLDAGNELPLTVYSRKRKPGFOHKKAGAMN 526
DB	479 MADGTAMPNGNPRDHPGMIOYFELHSGGLDIDGNELPLTVYSRKRKPGFOHKKAGAMN 538
QY	527 ALVRVSAVLTPGPFILINDCDHYINNSKALREAMCFIEMDNIGKQVYVQEPQRFQIDK 586
DB	539 ALIRVSAVLTPGAVLINDCDHYFNSSKALREAMCFIMDPALGRKYCYVQEPQRFQIDK 598
QY	587 NDRVANRRTVPEFDINLRGLDIOGVYVYGTCCVFNRFLVYEPRIYKHKRPSLSTLC 646
DB	599 HDRVANRRTVPEFDINLRGLDIOGVYVYGTCCVFNRQALVYGDPLTADLEPNLYIKSC 658
QY	647 GGSRRKSKAKKESKKKSGRHTDSTVVFMLDIEEVEGADGDEKALMLSQMSLEKR 706
DB	659 CGRRKKKKKSYMDSQSRIMKR--TESSADIFIMNEDIEEGE--GYEDERSVLMSSQKLEKR 715
QY	707 FGQSAVFAVASTLMENGYPSPATPEBNLKEALHYISCGYERKSDMGMTGMYISGVTEDI 766
DB	716 FGQSPISFIASFEMTQGGIIPSTNPASLKEALHYISCGYERKTEMGKEIGMYISGVTEDI 775
QY	767 LTGFRRMARGRSITYCMPKLPAPFGSAPAINLSDRLNOVLRYALAGSVELLFSRHCPITWGY 826
DB	776 LTGRKMARHGRQSTICMPPRCPFGSAPAINLSDRLNOVLRYALAGSVELLFSRHCPITWGY 835
QY	827 NGRLLFLERFAYVNTTIYPIITSIPLMYCTLLAYCLFTNOPIIQOISINIASIWFLSFLS 886
DB	836 NGRLLFLERLAYINTIYVPIITSIPLIACYCPALCILLNKFETIPEISNYAMGFILFAS 895
QY	887 IFAPGILLEMRSQVIGIDEMRNEDQFWYVGVSAHLFAFGQILLVYAGIDINNFYTSKAS 946
DB	896 IFAPGILLELRMSQVIGIEDMRNEDQFWYVGVSAHLFAFGQILLVYAGIDINNFYTSKAS 955
QY	947 DEDDDFAELVYEFKMTTLLIPTTLLIVLVGVVAVGSAVINSQVSMPLGKLEFFARV 1006

DB 956 DEDGFAELYFKMTSLIPPTVIVNIGVAGISVAINSGYSGMGLFCKLPFSIMV 1015  
QY 1007 IVHLVPLKGLMGRNRTPTIYVWMSVLLASIFSLMWRIDPFTSRVGPDL-ECGINC 1065  
DB 1016 ILHLVPLKGLMGRNRTPTIYVWMSVLLASIFSLMWRIDPFTSRVGPDL-ECGINC 1075  
RESULT 14  
ID AAY84110 standard; Protein; 1074 AA.  
XX AAY84110;  
AC AAY84110;  
DF 03-JUL-2000 (first entry)  
XX  
DE Amino acid sequence of a maize cellulose synthase.  
KM Maize; cellulose synthase; stalk quality; stand; silage; cellulose;  
KW transgenic plant; plant breeding marker.  
XX  
XX Zea mays.  
XX WO200009706-A2.  
XX 24-FEB-2000.  
XX 16-AUG-1999; 99WO-US18760.  
XX 17-AUG-1998; 98US-0096822.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Dhugga KS, Helentjaris TG, Bowen BA, Wang X:  
PI WPI: 2000-224343/19.  
DR N-PSDB: AA299500.  
XX  
XX New genes which encode maize cellulose synthase polypeptides in plants  
PT useful for modulating the expression of cellulose synthase in plants  
PT and to produce transgenic plants expressing the novel protein  
XX  
PS Claim 15; Page 110-113; 119pp; English.  
XX  
XX The present sequence represents a maize cellulose synthase polypeptide.  
CC The cellulose synthase can be used for the improvement of stalk quality  
CC for improved stand or silage. It also provides an increased concentration  
CC of cellulose in the pericarp, hardening the kernel and improving its  
CC handling ability. The sequences are used to produce transgenic plants  
CC and seeds expressing the cellulose synthase. The polynucleotide is  
CC used for modulating, preferably increasing, the level of the synthase  
CC in a plant cell. The plants are preferably monocots. The polynucleotide  
CC is also used as a probe or primer in the detection quantitation or  
CC isolation of gene transcripts. The probes are useful in detecting  
CC deficiencies in the level of mRNA in screenings for desired transgenic  
CC plant, for detecting mutations in the gene, for monitoring upregulation  
CC of expression or changes in enzyme activity in screening assays of  
CC compounds, for detection of any number of allelic variants of the gene,  
CC or for use as molecular markers in plant breeding programs. The  
CC isolated nucleic acids of the present invention can also be used for  
CC recombinant expression of their encoded polypeptides or for use as  
CC immunogens in the preparation and/or screening of antibodies. The  
CC proteins can be employed in assays for enzyme agonists or antagonists  
CC of enzyme function or for use of immunogens or antigens to obtain  
CC antibodies specifically immunoreactive with a protein.  
XX  
SQ Sequence 1074 AA;  
Query Match 70.9%; Score 4024; DB 21; Length 1074;  
Best Local Similarity 69.4%; Pred No. 0;  
Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;

DB 23 DGDADVPAKPTNSANGQVQICGDIYGVASATGDVFAACBECAPPCRCPEYERKRGNC 82  
QY 62 CPOCKTRKRLKSPAIPOCKDEGLADEGTYEFTNPOKEKISERMLGHHLTRGCEENG 121  
DB 83 CPOCKTRKRLKSPAIPOCKDEGLADEGTYEFTNPOKEKISERMLGHHLTRGCEENG 128  
QY 122 EPQYDKEYS-----HNHLPRLTSSKODTSGEFSAASPERLSVSTIAGCKRLPYSPDN 174  
DB 129 LOGDDADSSSARHPHRTIPRLTSGQISGETPDASPDHSTI-----RSPSTSYVD 180  
QY 175 QS--PNRRIYDP-----VGLGNVAMKERVDCKMKMKOEKNTGPVSTOASSEGGVDIDAS 226  
DB 181 PSVPVPRIVDPSKDLNSGLNSVDMKEVESMRVKODKNMLQVNTKRYEARGDME--G 237  
QY 227 TDILADEALLDNDAROPLSRKVSTPSSRIINPIMVIMLVLCLPFLHRTNPPVNAFA 286  
DB 238 TGSNGEDMQVMDARPLSRIVPISNOLNLRIVILRLILCEFFOYRISHPYRNAYG 297  
QY 287 LMLVSVICWIFALSMIIDOPPKWPPVNETYTLRLALYDRGEPSSOLAADIVSTVD 346  
DB 298 LMLVSVICWIFALSMIIDOPPKWPPVNETYTLRLALYDRGEPSSOLAADIVSTVD 357  
QY 347 PLKEPPLVANTVLSILAVDYPVDKVCYVPDGAAMLSFESLAETSEFARKMVPFCCKY 406  
DB 358 PLKEPPLVANTVLSILAVDYPVDKVCYVPDGAAMLSFESLAETSEFARKMVPFCCKY 417  
QY 407 STEPPAPEMFAKIDYLDKQVTSFVKDRAMKREYEEFKIRINALVSKALCKPEEGV 466  
DB 418 NIEPRAPEEYFAOKIDYLDKQVTSFVKDRAMKREYEEFKIRINALVSKALCKPEEGV 477  
QY 467 MODGTPMPPGNNTGDHPGMIQVFLGNGVGLDAGNELPRLYVYSREKRPFGOHKKAGAN 526  
DB 478 MADGTAMPGNRRDHPGMIQVFLGNGVGLDAGNELPRLYVYSREKRPFGOHKKAGAN 537  
QY 527 ALVRSAYVLTNGPFLTNDLDCDHYINNKKLRBAMCFMLDPNGLKQVCYVQFQOREGIDK 586  
DB 538 ALVRSAYVLTNGPFLTNDLDCDHYINNKKLRBAMCFMLDPNGLKQVCYVQFQOREGIDK 597  
QY 587 NDRYANRNTVFEDINLRGIDGIGQPYVGTGCVFNTALYGEPPKVNKKPSSLKSC 646  
DB 598 HDRYANRNTVFEDINLRGIDGIGQPYVGTGCVFNTALYGEPPKVNKKPSSLKSC 657  
QY 647 GGSRRKNSAKKESKSKSKSRHTDSTVPFNLDIDEEGAGFDEKALALSONSLER 706  
DB 658 GGSRRKNSAKKESKSKSKSRHTDSTVPFNLDIDEEGAGFDEKALALSONSLER 714  
QY 707 FGQSAVFVASTLMENGVPSPATPENTLKEAIVHISCGYEDKSDMGMEIGWYGVTEDI 766  
DB 715 FGQSPFLASTFTMGQGIPTNPASLKEAIVHISCGYEDKSDMGMEIGWYGVTEDI 774  
QY 767 LTGFMAHARGMSIYCMPLPAFKSAPINLSDRLQVLRALGSEVILFSRRCPTWY 826  
DB 775 LTGFMAHARGMSIYCMPLPAFKSAPINLSDRLQVLRALGSEVILFSRRCPTWY 834  
QY 827 NGRLEFLERFAVNTTYITISIPLMYCTLLAVCFITQFTIPOSNIASIMFLSFLTS 886  
DB 835 NGRLEFLERFAVNTTYITISIPLMYCTLLAVCFITQFTIPOSNIASIMFLSFLTS 894  
QY 887 IFATGILERMWSGVGIDEMWRNDQFVWIGTSVLAFAVPOGLLKVLAGIDTNTVYSKAS 946  
DB 895 IFATGILERMWSGVGIDEMWRNDQFVWIGTSVLAFAVPOGLLKVLAGIDTNTVYSKAS 954  
QY 947 DEDGFAELYFKMTSLIPPTVIVNIGVAGISVAINSGYSGMGLFCKLPFSIMV 1006  
DB 955 DEDGFAELYFKMTSLIPPTVIVNIGVAGISVAINSGYSGMGLFCKLPFSIMV 1014  
QY 1007 IVHLVPLKGLMGRNRTPTIYVWMSVLLASIFSLMWRIDPFTSRVGPDL-ECGINC 1065  
DB 1015 ILHLVPLKGLMGRNRTPTIYVWMSVLLASIFSLMWRIDPFTSRVGPDL-ECGINC 1074  
RESULT 15

AA84112  
ID AAY84112 standard; Protein; 1074 AA.  
XX  
AC AAY84112;  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE Amino acid sequence of a maize cellulose synthase.  
XX  
KW Maize; cellulose synthase; stalk quality; stand; silage; cellulose;  
transgenic plant; plant breeding marker.  
XX  
OS Zea mays.  
XX  
PN WO200009706-A2.  
XX  
PD 24-FEB-2000.  
XX  
PE 16-AUG-1999; 99MO-US18760.  
XX  
PR 17-AUG-1998; 98US-0096822.  
XX  
PA (PION-) PIONEER HI-BRED INT INC.  
XX  
PI Dhuga KS, Helentjaris TG, Bowen BA, Wang X;  
XX  
DR WPI: 2000-224343/19.  
XX  
DR N-PSDB; AA299506.  
XX  
PT New genes which encode maize cellulose synthase polypeptides in plants  
PT useful for modulating the expression of cellulose synthase in plants  
PT and to produce transgenic plants expressing the novel protein  
PS Claim 15; Page 126-128; 119pp; English.  
XX  
XX The present sequence represents a maize cellulose synthase polypeptide.  
CC The cellulose synthase can be used for the improvement of stalk quality  
CC for improved stand or silage. It also provides an increased concentration  
CC of cellulose in the pericarp, hardening the kernel and improving its  
CC handling ability. The sequences are used to produce transgenic plants  
CC and seeds expressing the cellulose synthase. The polynucleotide is  
CC used for modulating, preferably increasing, the level of the synthase  
CC in a plant cell. The plants are preferably monocots. The polynucleotide  
CC is also used as a probe or primer in the detection quantitation or  
CC isolation of gene transcripts. The probes are useful in detecting  
CC deficiencies in the level of mRNA in screenings for desired transgenic  
CC of expression or changes in enzyme activity in screening assays of  
CC compounds, for detection of any number of allelic variants of the gene,  
CC or for use as molecular markers in plant breeding programs. The  
CC isolated nucleic acids of the present invention can also be used for  
CC recombinant expression of their encoded polypeptides or for use as  
CC immunogens in the preparation and/or screening of antibodies. The  
CC proteins can be employed in assays for enzyme agonists or antagonists  
CC of enzyme function or for use of immunogens or antigens to obtain  
CC antibodies specifically immunoreactive with a protein.  
XX  
SQ Sequence 1074 AA;  
Query Match 70.9%; Score 4024; DB 21; Length 1074;  
Best Local Similarity 69.4%; Pred. No. 0;  
Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;

DB 129 LQGDADLSSAKHDPHRIPLRTSGQISGELIPDASPDRHSI-----NSPISSYVD 180  
QY 175 QS--PNNRIYDP-----VGLGNVANKERVDGKMKOEKNTGPTQASERGVDIDAS 226  
DB 181 PSVPVPRVRIYDPSKDLNSYGLNSVDMKEKVESRNVQDMQMLQYTKKYEARGDMF---G 237  
QY 227 TDILADEALLNDARPOLSRKVSIPSSRINPRMVTMLRLVILLCFLHRTINPNAPRA 286  
DB 238 TSGNGEDMQVDARPLSRIVDISNQLNRYIILILILCEFFQYRISHPNVNAV 297  
QY 287 LMLVSYICETWPLSMIILOFPKFWPNVNETYIDRLALYDGEPSQLAADIFSTVD 346  
DB 298 LMLVSYICEWPLSMIILOFPKFWPINNETYIDRLALYDGEPSQLAADIFSTVD 357  
QY 347 PLKEPPLTANTPLSTLAADVVDKVCYVFDGGAAMLSPESLAETSEFARKVPPCKRY 406  
DB 358 PLKEPPLITANTPLSTLAADVVDKVCYVFDGGAAMLSPESLAETSEFARKVPPCKRY 417  
QY 407 STEPRAPWYFAKIDYLDKQVTSFVKDRAMKREYEFKIRINALVSKALCPREGNV 466  
DB 418 NIEPRAPWYFAKIDYLDKQVTSFVKDRAMKREYEFKIRINALVSKALCPREGNV 477  
QY 467 MDQGTWPGNNTGDHGMIOVELGONGGIDABENELPRVLYYSREKRPFGHKKAGANN 526  
DB 478 MDGTWPGNNTGDHGMIOVELGONGGIDABENELPRVLYYSREKRPFGHKKAGANN 537  
QY 527 ALVRSVAVLTNGPFIINTLDCDHYINNSKALREAMECLMPDNLKQCYVQFOPRFGIDK 586  
DB 538 ALVRSVAVLTNGPFIINTLDCDHYINNSKALREAMECLMPDNLKQCYVQFOPRFGIDK 597  
QY 587 NDYRANRNTVPEDINLRGIDGLOGPVYVGTGCVFNFTALYGPPIKVRHKKPSLSKLC 646  
DB 598 HRYRANRNTVPEDINLRGIDGLOGPVYVGTGCVFNFTALYGPPIKVRHKKPSLSKLC 657  
QY 647 GSRKRNKAKKESDOKKGRHDTVPVFNIDIEGVEGAFODEKALNSOMSLER 706  
DB 658 GSRKRNKAKKESDOKKGRHDTVPVFNIDIEGVEGAFODEKALNSOMSLER 714  
QY 707 FGQSAFVASTLMENGVPSPATPENILKEAIVHISGVEDKSDMGMEIGWYGVTEDI 766  
DB 715 FGQSPFIASTFTMGQIPSPINPALKEAIVHISGVEDKSDMGMEIGWYGVTEDI 774  
QY 767 LTFGKHANRMSIYCMPLPAFKGAPINLSDRNLQVRLMALGSVEILLSHRCPIWGY 826  
DB 775 LTFGKHANRMSIYCMPLPAFKGAPINLSDRNLQVRLMALGSVEILLSHRCPIWGY 834  
QY 827 NGRKLEPRFAYVNTIYPTISPILMYCTLLAVCLFTNOIPTSINIASIWPSTLFTS 886  
DB 835 NGRKLEPRFAYVNTIYPTISPILMYCTLLAVCLFTNOIPTSINIASIWPSTLFTS 894  
QY 887 IFATGILEMRWSGVGIDEMWRNDEQFVIGVSAHLEFAVQGLKVLADIDNFTVSKAS 946  
DB 895 IFATGILEMRWSGVGIDEMWRNDEQFVIGVSAHLEFAVQGLKVLADIDNFTVSKAS 954  
QY 947 DEDGDFAEIYLFKWTLLIPPTLLVNLGVAVAGVAINSGYOSMGPLFGKLPFAFNV 1006  
DB 955 DEDGDFAEIYLFKWTLLIPPTLLVNLGVAVAGVAINSGYOSMGPLFGKLPFAFNV 1014  
QY 1007 IYHLVPELKLGMGRORPPTIVVMSVLASTFSLIMVITDPTSVTPDIL-DEGINC 1065  
DB 1015 IYHLVPELKLGMGRORPPTIVVMSVLASTFSLIMVITDPTSVTPDIL-DEGINC 1074

Search completed: June 16, 2003, 11:06:41  
Job time : 50 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:03:13 ; Search time 15 Seconds

(without alignments)  
2944.819 Million cell updates/sec

Title: AAC39336

5677

Sequence: 1 MESEGETAGKPMKNIVPQC.....IDPFTSRVGPDIIECGINC 1065

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273	4.8	872	1	BCSA_ECO57
2	273	4.8	874	1	BCSA_SALTI
3	272	4.8	872	1	BCSA_ECOLI
4	271	4.8	874	1	BCSA_SALTY
5	253.5	4.4	729	1	BCSA_XANAC
6	253.5	4.4	1596	1	ACS2_ACCEX
7	247.5	4.4	1518	1	BCA4_ACCEX
8	247.5	4.4	1518	1	BCA5_ACCEX
9	236.5	4.2	756	1	BCA2_ACCEX
10	228.5	4.0	754	1	BCA1_ACCEX
11	228	4.0	739	1	BCSA_PSEFL
12	208.5	3.7	745	1	BCA3_PSEFL
13	196	3.5	1550	1	ACS1_ACCEX
14	119	2.1	517	1	COX1_PARLI
15	118.5	2.1	590	1	YKTA_CAEL
16	115	2.0	357	1	G6PT_HUMAN
17	115	2.0	517	1	COX1_ASTPE
18	110.5	1.9	441	1	YCDQ_ECOLI
19	110.5	1.9	672	1	ALIB_HUMAN
20	110	1.9	517	1	COX1_STRPU
21	109	1.9	830	1	FAR1_YEAST
22	108	1.9	517	1	COX1_PISOC
23	108	1.9	1029	1	RIP3_RAT
24	107.5	1.9	1437	1	MRP5_HUMAN
25	107	1.9	3033	1	POLG_HCV6
26	106	1.9	357	1	G6PT_CANFA
27	105.5	1.9	1076	1	YE38_SCHPO
28	105.5	1.9	4829	1	BIR6_HUMAN
29	104.5	1.8	309	1	YESP_BACSU
30	104	1.8	995	1	AHM7_ARATH
31	104	1.8	1167	1	CIRA_BACTU
32	104	1.8	1879	1	POLN_SMSV1
33	103	1.8	414	1	YOAB_BACSU

34	103	1.8	470	1	ROCC_BACSU
35	103	1.8	1066	1	NUC2_NEUCR
36	103	1.8	1232	1	B3A3_HUMAN
37	102	1.8	642	1	ARE2_YEAST
38	101.5	1.8	1863	1	BRC1_HUMAN
39	101	1.8	617	1	VAA1_BOVIN
40	101	1.8	3097	1	CADN_DROME
41	100.5	1.8	744	1	YTV2_CAEL
42	100.5	1.8	1020	1	ATC1_DROME
43	99.5	1.8	322	1	NUIM_STRPU
44	99.5	1.8	536	1	COX1_ALUMA
45	99.5	1.8	714	1	HUNK_MOUSE

## ALIGNMENTS

RESULT 1  
BCSA\_ECO57  
ID BCSA\_ECO57 STANDARD: PRT: 872 AA.  
AC O8XSL7;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).  
GN BCSA OR 24948 OR EC54413.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDE933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackert J., Evans P.S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamosis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RT Nature 409:529-533(2001).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11238796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12.";  
RT DNA Res. 8:11-22(2001).  
-1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes  
uridine 5'-diphosphate glucose to cellulose, which is produced as  
an extracellular component for mechanical and chemical protection  
at the onset of the stationary phase, when the cells exhibit  
multicellular behavior (rod morphology). Co-expression of  
cellulose and thin aggregative fimbriae leads to a hydrophobic  
network with tightly packed cells embedded in a highly inert  
matrix (By similarity).  
-1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))<sub>n</sub> = UDP  
+ ((1,4-beta-D-glucosyl))<sub>n+1</sub>.  
-1- CORRECTOR: Magnesium (By similarity).  
-1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid  
(C-di-GMP) (By similarity).  
-1- PATHWAY: Bacterial cellulose biosynthesis.  
-1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
(Potential).  
-1- DOMAIN: There are two conserved domains in the globular part of  
the protein: the N-terminal domain (domain A) contains  
the conserved DXD motif and is possibly involved in catalysis and  
substrate binding. The C-terminal domain (domain B) contains the



CC substrate binding. The C-terminal domain (domain B) contains the  
CC QXXRW motif and is present only in processive glycosyl  
CC transferases. It could be involved in the processivity function of  
CC the enzyme, possibly required for holding the growing glycan chain  
CC in the active site.  
CC  
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
CC  
CC EMBL; AL627281; CAD08006.1; -  
CC InterPro; IPR001173; Glycosyltransf\_2.  
CC Pfam; PF00535; Glycos. transf. 2; 1.  
CC KW Cellulose holoenzyme; Transferrase; Glycosyltransferase;  
CC Transmembrane; Inner membrane; Complete proteome.  
CC FT DOMAIN 271 364 CATALYTIC SUBDOMAIN A.  
CC FT TRANSMEM 30 50 CATALYTIC SUBDOMAIN B.  
CC FT TRANSMEM 151 171 POTENTIAL.  
CC FT TRANSMEM 173 193 POTENTIAL.  
CC FT TRANSMEM 230 250 POTENTIAL.  
CC FT TRANSMEM 525 545 POTENTIAL.  
CC FT TRANSMEM 547 567 POTENTIAL.  
CC FT TRANSMEM 592 612 POTENTIAL.  
CC FT TRANSMEM 634 654 POTENTIAL.  
CC FT TRANSMEM 668 688 POTENTIAL.  
CC FT TRANSMEM 833 853 POTENTIAL.  
CC FT ACT\_SITE 313 313 POTENTIAL.  
CC FT ACT\_SITE 457 457 POTENTIAL.  
CC FT SITE 360 360 SUBSTRATE BINDING (POTENTIAL).  
CC FT SITE 362 362 SUBSTRATE BINDING (POTENTIAL).  
CC SQ SEQUENCE 874 AA; 100020 MW; 3395D09CD051990 CRC64;  
CC  
CC Query Match 4.8%; Score 273; DB 1; Length 874;  
CC Best Local Similarity 18.8%; Pred. No. 8.6e-12;  
CC Matches 153; Conservative 91; Mismatches 201; Indels 370; Gaps 29;  
CC  
CC QY 260 MVIVLRLVILICLFHYRTITNV---PRAFLMVLVSVICELIFALSWILDDPPKKFPVNR 315  
CC Db 201 MLVILSVSCRYTWMRTSTILNDDPVSLVCGILFAETRYAMVLVLGYFOVWPLNR 260  
CC QY 316 ETVLDRLALRDREGSPQLAOLVDFSTVDRLEKRPVLTNTVSLIANDVPDYKASY 375  
CC Db 261 QP-----VPLREKNSQMTVDIEFVTYN---EDLVNKKNTIASIGIDMPKCKLNIW 309  
CC QY 376 VFDDGAAMLSFESLAETSEFARKVVRPECKKYSIEPRAEWYFAAKIDYLDKDKVQTSFVKD 435  
CC Db 310 ILDDGGR---ESF---RQFAR----- 324  
CC QY 436 RRANKREYEEKIKINALVSKALCKREGGWMODSTPWPAGNITGDHPMIOVFLGONGSL 495  
CC Db 325 -----HVGYN----- 329  
CC QY 496 DAEGNELRLVYVREKRPQFQNHKAGAMNALRVSAVLNGLPILMLDDCHYINNSKA 555  
CC Db 330 -----YLR---ATHEHAKAGINNALNHA---KGEVALFDCHDVTRSF 370  
CC QY 556 LREAMEGLMDPRLKQOVYVQFQRF---DGIDKN---DRYANRETVFEDINLRGLDGI 608  
CC Db 371 QMTGWGFLKE---KQLAMQOTRNHFERRDPERENLRGFRKTRPNNGCLTFYGLVDGNDMM 426  
CC QY 609 QGPRVYVTCGYFNRTALVGLYERPLTKVNHKKPSLSLKCGSKSKKSKAKKESDKKSGRH 668  
CC Db 427 DATEFGSCCAVIR-----KP----- 442  
CC QY 669 TDSVVRVFNLDIDIEGVAGAGFDEKALLMSQMSLEKRFQGSANVAVASTLMENGVPVPSA 728  
CC Db 443 -----LDEI-----GGI----- 449

Oy		729	TPEMLLEAETHAVICSGYEDKSDMKMEIGWITGSVTEDLTIDLPFKMNAHRCMBSTGYCPRKPA	788
Dd		450	-----AVE-----       : :       :   :	478
Oy		789	EKSQAPINLSORLNQVRLMALGVSVEILFSRHCPIMWGNGRLKELERFAYVNTTYPTTS	848
Dd		479	SAGLATESLSAHIOGRIRRMANGVOI-TRDNLPL---FGKRLAKLAQRCLYNAMHFLLSG	534
Oy		849	IPLIMYCT-----LLAVCLEFTNOFIPOI-----SNIASIFLSEFLSIFA	889
Dd		535	IPRILFETAPLAFELLHAIIYYAPALMALIALVFIPHMVASLTNSIKOGKYHSHSEIVE	594
Oy		890	TGLEMRNRSGCIDEMWRNEGFWMYGISALFAVFOGLVLVLAGIDINFMTSKASSED	949
Dd		595	TVLA-----WTYAPT-----LVALLNDHKRENVTK-----	622
Oy		950	GDFAEELYLFKWTTLPIPTTLLIYNLGVAGVSAYAINSOSMGPLEPAAFFFWVIWH	1009
Dd		623	GGIVEEKYVDN-VYSRPYIFVLINTLLGVAGV-----WRYYPG-----	661
Oy		1010	LYPEKLGMGRONRTPIIVVVWSVLLASIFSCLAW	1044
Dd		662	-----ENETLTVIV-----SLVWV	675
<b>RESULT 3</b>				
BCSA_ECOLI				
ID	BCSA_ECOLI	STANDARD;	PRT;	872 AA.
AC	P37653; P37654; P76712; P76713;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).			
GN	BCSA OR B3533.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
CC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / MG1655;			
RX	MEDLINE=94316500; Pubmed=8041620;			
RA	Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;			
RT	"Analysis of the Escherichia coli genome. V. DNA sequence of the			
RL	region from 76.0 to 81.5 minutes."			
RN	Nucleic Acids Res. 22:2576-2586(1994).			
RN	[2]			
RP	REVSIONS.			
RC	STRAIN-K12 / MG1655;			
RX	MEDLINE=97426617; Pubmed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley J.M., Collado-VIDes J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.;"			
RL	Science 277:1453-1474(1997).			
RN	[3]			
RP	CHARACTERIZATION			
RC	STRAIN-ECCORI2, ECCORI2, and TOB1;			
RX	Pubmed=11260463;			
RA	Zozag X., Nimtz M., Rohde M., Bokranz W., Roemling U.;			
RT	"The multicellular morphotypes of Salmonella typhimurium and			
RT	Escherichia coli produce cellulose as the second component of the			
RT	extracellular matrix."			
RL	Mol. Microbiol. 39:1452-1463(2001)."			
CC	-1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes			
CC	uridine 5'-diphosphate glucose to cellulose, which is produced as			
CC	an extracellular component for mechanical and chemical protection			
CC	at the onset of the stationary phase, when the cells exhibit			
CC	multicellular behavior (rilar morphology). Co-expression of			
CC	cellulose and thin aggregative fimbriae leads to a hydrophobic			
CC	network with tightly packed cells embedded in a highly inert			

RESULT 3  
BCSA\_ECOLI STANDARD: PRT: 872 AA.  
ID BCSA\_ECOLI P37653; P37654; P76712; P76713;  
AC P37653; P37654; P76712; P76713;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).  
GN BCSA OR B3533.  
OS *Escherichia coli*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC *Escherichia*.  
OX NCBI\_TaxID=562;  
RN [1]  
RP  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=94316500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT "Analysis of the *Escherichia coli* genome. V. DNA sequence of the  
RL region from 76.0 to 81.5 minutes.";  
RN Nucleic Acids Res. 22:2576-2586(1994).  
[2]  
RP  
RP REVISIONS  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.R., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RT Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
[3]  
RN  
RP CHARACTERIZATION:  
RC STRAIN=ECOR10, ECOR12, and TOB1;  
RX PubMed11260463;  
RA Zogaj X., Nlantz M., Rohde M., Bokranz W., Roemling U.;  
RT "The multicellular morphotypes of *Salmonella typhimurium* and  
RT *Escherichia coli* produce cellulose as the second component of the  
RT extracellular matrix";  
RL Mol. Microbiol. 39:1452-1463(2001).  
-1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes  
uridine 5'-diphosphate glucose to cellulose, which is produced as  
an extracellular component for mechanical and chemical protection  
at the onset of the stationary phase, when the cells exhibit  
multicellular behavior (rod morphology). Co-expression of  
cellulose and thin aggregative fimbriae leads to a hydrophobic  
network with tightly packed cells embedded in a highly inert

Db	310	ILDDGG-----	315
Oy	436	RRAKREYEERKIRINLALVSKALCPREGWVMOGSTPRGNNTGDPHMIOVYIQLGNGL	495
Db	316	---REEROP-----	328
Oy	496	DAEGNELPRLYVVSREKRPFGOHKKAGAMALVRSVAVLTNGPILNDCDHYINNSKA	555
Db	329	K-----YIARTT-----HEHAKAGINNALKRYA-----KGEVSTLFEDDHYPTBRFL	370
Oy	556	LREAWCEFLMDNPLGKVCYVOPORF---DGIDKN---DRYANRNVFEDI NLRGLDGI	608
Db	371	QMTGMWFLKE---KOLAMQOTPRHFESPDPEERNLGRFRRTPNEGTFLGYLMDGNDMM	426
Oy	609	QGPVYVVGCGFENFTALGYEPRLKVKKKPSLLSKLGGSKKKKKSKKSSDKKSGRH	668
Db	427	DATEFCGSCAVIRR-----KP-----	442
Oy	669	TDSTVYPVFNLDIEGVEGAGFDEDEKALLMSOMSLERKRGQSAVFAVASTMBNGVPPSA	728
Db	443	-----LDEI-----	449
Oy	729	TPEMLKEAHIVISCGYEDKDKMGMEIGMIYGSVTEOILTFGPKHARGRSITPCMKPLA	788
Db	450	-----AVE-----	478
Oy	789	FKSGAPINLSRLNOVLFWALGSEVLEFRRCPIWYGNGR-LKFLERFAVNTTIYPT	847
Db	479	AAGLATELSAHIGORIRMARWGVOI-FRLDNPL---TGKGLKPAQGLCYVNMHFLS	533
Oy	848	SIPLLMCTLLAVCLFTNQFIIPQISINIASIWFSLFESTATGILEKMSGVGIDEMWR	907
Db	534	GIPRIETLAPLAEFLLAHYIT---YAPALMALFV---LPHNIRHASLNSKIOGKYR	585
Oy	908	NEOFNVAGVSANLFAVPOGIL-----KVLAGIDTN---FTYTSKASDEGPAELYL	957
Db	586	H-SFW-----SEIETVLANITYAPPTVALINPKKGFNTAK-----GGLVEEY	630
Oy	958	EKWTTLLIPTTLLIYNLVGVAGVSAINSYGSGWGLFGKLFFAFVIVHLYPELKL	1017
Db	631	VDM-VISPRYFVLVNLGVAVGI-----WRFYFG-----	660
Oy	1018	MGRONRPTIYVWVSVLASIFSLTW 1044	
Db	661	-----PPT-----EMLTVVSMWV 675	
RESULT 4			
BCSA_SALTY			
ID	BCSA_SALTY	STANDARD:	PRT: 874 AA.
AC	O9JIN2;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Cellulose synthase catalytic subunit (UDP-forming) (EC 2.4.1.12).		
GN	BCSA OR SWM3619		
OS	(Salmonella typhimurium)		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
CC	Salmonella.		
RN	NCBI_TaxID=602;		
RN	111		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 14028;		
RK	MEDLINE=21160181; PubMed=11260463;		
RA	Zogaj X., Nintz M., Rohde M., Boktranz W., Roemling U.;		
RT	"The multicellular morphotypes of Salmonella typhimurium and		
RT	Escherichia coli produce cellulose as the second component of the		
RT	extracellular matrix."		
RL	Mol. Microbiol. 39:1452-1463(2001).		
RN	121		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=LT2;		
RK	MEDLINE=21927388; PubMed=11929533;		

FT	ACT_SITE	313	313	POTENTIAL.
FT	ACT_SITE	457	457	POTENTIAL.
FT	SITE	360	360	SUBSTRATE BINDING (POTENTIAL).
FT	SITE	362	362	SUBSTRATE BINDING (POTENTIAL).
SO	SEQUENCE	874 AA;	100043 MW;	4C9421B58606310A CRC64;
Query Match				
Best Local Similarity		4.88;	Score 271;	DB 1; Length 874;
Matches 153; Conservative		18.8%; Pred. No. 1.2e-11;	90; Mismatches 202; Indels 370; Gaps 29;	
OY	260	MVIMLRVILLCFLHYRTINPV----	PNAFLMLVSYICEIWFPLSNILDOFPKMPFVNR	315
DB	201	MLIYLSLTVSCRIYIMWRITSLIMDDPVSILYCGILLFAEYTAIVTLVIGFYQWVPLNR	260	
OY	316	ETYLDRALRYRDEGERSOLAANDIFVSTVDPLKEPPLVANTVLSILANDPVDKVSCY	375	
DB	261	QP-----VPLRKESQMPFTVIDIFVPTVN---	EDLVNVTAVISLIGIDMPKDKLINW	309
OY	376	VFDGAMLSSESLAETSEFARKKVPCKKTSIEPRAPEWIFAAKIDLKDKVTSFYKD	435	
DB	310	ILDDGGR---ESF---RHFAK---		324
OY	436	RRAMKREYEEFKINMLVSKALCPRBGWVWDQSTPWPNGNTDHDHGMIGVFGONGL	495	
DB	325	-----	HGVH-----	329
OY	496	DAEGNELRLVYVSREKRPFOHKKKAGAMNALVRSVAVLTNGPFIINLDCDHYNNSKA	555	
DB	330	-----YIARTT---HEHAKGNINNALKHA---	KGEVAFIDCDHVPTRSF	370
OY	556	LREAMCFIMDPNLKQVCYVOPRPF---DGIDKN---	DRIANRNVFPIINRGIDGI	608
DB	371	QMTGWFLKE---KOLAMQTPRHFFSPDFEENLGRFRTPREGTLEFGLVODGDNDMW	426	
OY	609	QGVYVVGCGCFENFTALGYEPRPIKVKKKKSLSLKLCGGRKKNSKKKKSKKKSGRH	668	
DB	427	DATEFCGCAVIR-----	KP-----	442
OY	669	TDSTVVPFNLDIEGEGAGFDEKALIMSQMSLEKFGQSAVFVASTLMENGWPPSA	728	
DB	443	-----LDEI-----	-----GGI-----	449
OY	729	TPENILKEAHIVISCGYEDKSDMGMEIGWIGSVETEDILGFKMHARGMSIYCMPLRA	788	
DB	450	-----AVE-----	-----TVTEADHTSLRHRRGYTSAYM--RIPQ	478
OY	789	FKGSAPILSRLNOVLRNALGSEVILEFSRCRPIWYGVNGNLKFLERAVYNTTIPYPTS	848	
DB	479	AAGLATESLSAHIOIRIMARGWMOI--FRIDNPL--	FGKGLKLAORLCYLIAMHNFHISG	534
OY	849	IFPLMYC-----	LLAVCLFNQFIPOI-----	889
DB	535	IPRLIFLTAPLAFLLAHYIIYALMALAFVIFHMVHASLTSNKIOGKHNSRSEIYE	594	
OY	890	TGILEMRKSGVIDEMWRNEQFWYIGVSAHLFAVFOGILVLVLAGIDTNTFTYKASQSD	949	
DB	595	TVLA-----	WTIAPPT-----	622
OY	950	GDFAEFLYFKTTLIPPTLLIYNLVNGVYAGVSYALNSGYSQSGPRLGALFFAPWYLVH	1009	
DB	623	GGIVEEKVVDN--VISRPYIFLVLLNLGLGVAAG-----	WRYVGG-----	661
OY	1010	LYPFLKGLMGHGRNRPRTIVVWVSVVLASIFSILWV	1044	
DB	662	-----ENETLVIYV-----	SLWV	675
RESULT 5				
AC	BCSA_XANAC	STANDARD:	PRT:	729 AA.
AC	P58932:			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			

FT	ACT SITE	337	337	POTENTIAL.
FT	SITE	240	240	SUBSTRATE BINDING (POTENTIAL).
FT	SITE	242	242	SUBSTRATE BINDING (POTENTIAL).
SO	SEQUENCE	729 AA:	80915 MW;	B9C08BB95E795B1 CRC64:
	Query Match	4.4%;	Score 252.5;	DB 1; Length 729;
	Best Local Similarity	19.6%;	Pred. No. 2e-10;	
	Matches	153;	Conservative	85; Mismatches 212; Indels 331; Gaps 28;
OY	260	WVIMRLVILLCFLHYRTINP--VENA--PALMLVSYCEIWFALSMILDDGFKKFPVNR	315	
DB	81	MMGMSLAVSCRYIMRMTOQMGVSADFLIGLLGALGAFAFVILVGFYGLPMPNR	140	
OY	316	ETVLDRLARVDRREGEPQSLAAVDLFVSTVPLKEPPLVTANTVLSILAVYFVDKVCY	375	
DB	141	KP----VPLPADQRLMP-----VDFFITYN---EPLSVRTYLAASVIDMPAKKIITH	189	
OY	376	VFDGAAMLSFESLAETSEFARKVVPCKKYSIEBRADEWYFAAKIDYLDKQVQTSFYKD	435	
DB	190	LDDG-----	194	
OY	436	RRAKRKEEERKIRINALVSKALCKPREGGWMQDSTPWRGNNTGHPMIGVFLQNGSL	495	
DB	195	--RDEERAF-----CAEYG-----	207	
OY	496	DAEGNELRLVYVSEKRPFOHKKAGAMNALVFSAVLTNGPEFLNDDCHYINNSKA	555	
DB	208	-----INVYRTNNA-----HAKAGNINAAK-----KSGDVAALFDDCH-IPTRSF	249	
OY	556	LREAN-CELPMPNIGKQYCYQFQRF-----DGIDKNDRYANRNVFEDLNLRGIDG	607	
DB	250	LQVAMGWLHDTKL--ALVQMPHYEFSPPDEFERNIDTHCKVPEGELFLGLDDGDNDQ	305	
OY	608	IQGPVYVGTGCVFNFTALYGYEPRPKYVKKRPSLLSKLCGSKRKKSKAKKESDCKKSGR	667	
DB	306	WNAFFEGSCAVIKRTA-----	322	
OY	668	HTDSTVPVNLDDIEEGVEGAGFDDKALMLMSLSLEKRFQSGAVFVASTLMENGGVPPS	727	
DB	323	-----LEEVGV-----	329	
OY	728	ATPENILKEAIVHIVISCGYEDKSDMGMEIGWLYGVSATEDILGFKMHARGMSIYCMPLP	787	
DB	330	-----AVE-----TYTEAHNALKIQRRGYKRAYL--AVP	357	
OY	788	AFKGSAPINLSDRLNQLRMALGSVEI-----LESRHCPIYWGNGRLKFLERFAYVNT	841	
DB	358	QAAGIATSESTSGHVAQRIRMARGMQAIRINDPILGR-----GLKLSQRLCYUNA	407	
OY	842	TIYPTISPLMLMCTCLLAVCLFTNQFIIPQISNLASIMFLSLFSIFATGI LEMRMESGV	900	
DB	408	MLHFHYGPRIITVLAFLAVLFEFGAHVYQASNLMLAVALPHILQANLTNLNVOSRFNHL	467	
OY	901	GIDEMWRNEQWVIGVSAHLEFAVFOGILYLAGIDT-NFTVYFSA-----SDEGDEPAE	954	
DB	468	LMNEVEYETTLAMYI-----FRPLVALLNPKICKFNNVTPGGVLAARYFPAQALAK	517	
OY	955	LYLKEWITLLIPPTLLILVNLGVYAGV---SYAINSGYQS--WCPFLGLLFFAFWVLYH	1005	
DB	518	PYLF-----LILLNVGMVAGVLRILYVSGSEQDTIW-----FNLAWTLTYN	559	
OY	1010	L 1010		
DB	560	M 560		
RESULT 6				
ACSG2_ACEXY				
ID	ACSG2_ACEXY	STANDARD;	PRT;	1596 AA.
AC	059167:			
DT	15-JUN-2002	(Rel. 41, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		

DE cellulose synthase 2 [includes: Cellulose synthase catalytic subunit  
DE (UDP-forming) (RC 2.4.1.12); Cyclic di-GMP binding domain (cellulose  
DE synthase 2 regulatory domain)].  
DN ACSA11.  
OS Acetobacter xylinus.  
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;  
OC Glucosaccharobacter.  
OX NCBI\_TaxID=28448;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 23769;  
RX MEDLINE=95394846; PubMed=766515;  
RA Saxena I.M., Brown R.M. Jr.;  
RT "Identification of a second cellulose synthase gene (acsAII) in  
RT Acetobacter xylinum.";  
RL J. Bacteriol. 177:5276-5283(1995).  
CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))<sub>(N+)</sub> = UDP  
CC + ((1,4-beta-D-glucosyl))<sub>(N+)</sub>.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (potential).  
CC -1- DOMAIN: There are two conserved domains in the globular part of  
CC the catalytic subunit: the N-terminal domain (domain A) contains  
CC the conserved DXD motif and is possibly involved in catalysis and  
CC substrate binding. The C-terminal domain (domain B) contains the  
CC OXXXW motif and is present only in processive glycosyl  
CC transferases. It could be involved in the processivity function of  
CC the enzyme, possibly required for holding the growing glycan chain  
CC in the active site.  
CC -1- MISCELLANEOUS: It is not essential for cellulose production in  
CC this strain.  
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
CC GLYCOSYLTRANSFERASE FAMILY 2.  
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSB/BCSB  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
CC or send an email to [license@slb.ch](mailto:license@slb.ch)).  
CC -----  
DR EMBL; U15957; AAA85264.1; -  
DR InterPro; IPR003919; Cellsynth\_A.  
DR InterPro; IPR003920; Cellsynth\_B.  
DR InterPro; IPR001173; Glycos\_transf\_2.  
DR Pfam; PF031170; BcsB; 1  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
DR PRINTS; PR01439; CELLSTHASE.  
DR PRINTS; PR01440; CELLSTHASEB.  
KW Cellulose biosynthesis; Transferase; Glycosyltransferase;  
KW Transmembrane; Inner membrane.  
FT DOMAIN 1 749 CATALYTIC.  
FT DOMAIN 750 1596 CYCLIC DI-GMP BINDING (POTENTIAL).  
FT DOMAIN 145 238 CATALYTIC SUBDOMAIN A.  
FT DOMAIN 315 375 CATALYTIC SUBDOMAIN B.  
FT TRANSMEM 25 45 POTENTIAL.  
FT TRANSMEM 106 126 POTENTIAL.  
FT TRANSMEM 396 416 POTENTIAL.  
FT TRANSMEM 421 441 POTENTIAL.  
FT TRANSMEM 505 525 POTENTIAL.  
FT TRANSMEM 544 564 POTENTIAL.  
FT TRANSMEM 1553 1573 POTENTIAL.  
FT ACT\_SITE 187 187 POTENTIAL.  
FT ACT\_SITE 331 331 POTENTIAL.  
FT SITE 234 234 SUBSTRATE BINDING (POTENTIAL).  
FT SITE 236 236 SUBSTRATE BINDING (POTENTIAL).  
SQ SEQUENCE 1596 AA; 175799 MW; D98A6F6259E1F3CE CRC64;  
Query Match 4.48; Score 252.5; DB 1; Length 1596;  
Best Local Similarity 17.78; Pred. No. 6.6e-10;  
Matches 142; Conservative 99; Mismatches 215; Indels 345; Gaps 25;

QY	261	VIMLVILVILCFLHRYRT-----NPRDAFALMVLS-----VICIWFALSMILDOF	307
Db	76	LMLSLSLRYMYWRLLTTLIELHSPDLAALSLILVAEYALALTLCISLYQMSHPDRK	135
QY	308	PKWEPVNRETYLDRLALRYDREGSPQLAAVDIEVSTVDPRLKPEPLVTANTVLSIAVDY	367
Db	136	PLPLP-----ADPTDMPVVVDVYPSVN---EELSLVRSYTVGLALAIWD	175
QY	368	PVDFVSCYVDFDGAAMLSFESLAETSEPAKRWPFCKKYSLEPAAEWYFAAKIDYLDKQ	427
Db	176	PADRLNVIYIIDDG-----188	
QY	428	VQTSFVNDRAKMKREVEEFKIRIINALYSKALKSREEGVMODGTPWPCNNTGDHPMTQV	487
Db	189	-----RRSSFAFAEAGA-----GIITND-----208	
QY	488	FLAGNGGLDAEGNELPLRVVVSREKRRGFQNHKKAGAMNALVRYSAVLTNGPFTLNLDCD	547
Db	209	-----GN-----NHAKAGNLNHALRV---TEGEVYVIFCD	236
QY	548	HYNNSKALREAMCFEL-DEPLGKQVCTVQRPQRF---DGIDKN---DRIANNVYFCD	599
Db	237	H-IPTRGFLKKTIGMMADPKL---ALDQPHNFYSPDPQRNLATGQNVPRGGMFYG	291
QY	600	INLGGLOGIQGVVYNGCVENPNTALGXEDRPYKVKHKPSLSKLGCSRKKSAAKE	659
Db	292	LVQDGNDFWDAITFCGSCAALRRSAVLGI-----320	
QY	660	SDKKSKGRHDTSTVPVFNLDIEEGVGAGFDEDEKALLMSQMSLEKRGQSAVFVAATLM	719
Db	321	-----GGFATE-----326	
QY	720	ENGVPFSAPTENLKEAIVHISCGYEEDKSDMGHEIGWYGSVEDILTGKMAKRGWS	779
Db	327	-----TVTEDANTALKMREGAHT	345
QY	780	IYCPMKLPAFKGSAPINLSDRLNOLRWALGSVELTFSRHCPIWYNGRGLFERAFVY	839
Db	346	AYLQPLPAA--GLSTERMLHIGQVRKARAKMLDMLRDNLLSG-----LRMOORLCYL	399
QY	840	NTTYPTITSIPLMYCTLLAVCLFTNOFTIPOISINASIW-FLSLFISFATGILEMWS	898
Db	400	SAMSHFLFAPRLVFLVSLPLFLFLGQNIIAASFALIVYAFPHVHISIGTLRSVEGRW-	458
QY	899	GVGIDEMKRNQEF-VIGCVSAHLFPAPQGLKVLGADTNTFTYAKSADSDGCPAEYL	957
Db	459	-----RYSEFSEIYETTLALFLVAVTITMLNPRKGFENVTQKGLLOSEFDL--	507
QY	958	FKWTTLTLPRTLLTILVNLGVAVV-----SYAINSGYQSMGPLCKLFFFA	1003
Db	508	-----NAVYPNVIILVIALALVLRIGCGMMMEYHRLALQSFALMT-----548	
QY	1004	FWIVVHLVPLKGL-MGRQNR	1023
Db	549	LMVAVSLIIVLASIYAVGRETR	569
RESULT 7			
BCA4_ACCEXY STANDARD; PRT; 1518 AA.			
AC Q9RBJ2: 15-JUN-2002 (Rel. 41, Created)			
DT 15-JUN-2002 (Rel. 41, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DE Putative cellulose synthase 2 [includes: Cellulose synthase			
DE catalytic subunit [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding			
DE domain (cellulose synthase 2 regulatory subunit)].			
GN BCSABIT-A.			
OS Acetobacter xylinus.			
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;			
OC Glucomacetobacter.			
OX NCBI_TaxID=28448;			

RN [1]  
 RP .SEQUENCE FROM N.A.  
 RC STRAIN-JCM 7664 / IFO 13693;  
 RX PubMed=10382968;  
 RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,  
 RA Inoue Y.,  
 RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM  
 RT 7664: implication of a novel set of cellulose synthase genes";  
 RL DNA Res. 6:109-115(1999).  
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))(N) = UDP  
 CC + ((1,4-beta-D-glucosyl))(N+1)  
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -1- DOMAIN: There are two conserved domains in the globular part of  
 CC the catalytic subunit: the N-terminal domain (domain A) contains  
 CC the conserved DXD motif and is possibly involved in catalysis and  
 CC substrate binding. The C-terminal domain (domain B) contains the  
 CC QXXRW motif and is present only in processive glycosyl  
 CC transferases. It could be involved in the processivity function of  
 CC the enzyme, possibly required for holding the growing glycan chain  
 CC in the active site.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
 CC GLYCOSYLTRANSFERASE FAMILY 2.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ACSE/BCSB  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AB015803; BAA77593.1; -  
 DR InterPro: IPR003919; CellSynth\_A.  
 DR InterPro: IPR001173; Glycosyl\_transf\_2.  
 DR Pfam: PF00535; Glycosyl\_transf\_2; 1.  
 DR PRINTS: PR01439; CELSINTHASE.  
 DR PRINTS: PR01440; CELSINTHASE.  
 KM Cellulose biosynthesis; Transferase; Glycosyltransferase;  
 KM Transmembrane; Inner membrane.  
 FT DOMAIN 1 731 CATALYTIC.  
 FT DOMAIN 732 1518 CYCLIC DI-GMP BINDING (BY SIMILARITY).  
 FT DOMAIN 144 237 CATALYTIC SUBDOMAIN A.  
 FT DOMAIN 314 374 CATALYTIC SUBDOMAIN B.  
 FT TRANSMEM 24 44 POTENTIAL.  
 FT TRANSMEM 71 91 POTENTIAL.  
 FT TRANSMEM 105 125 POTENTIAL.  
 FT TRANSMEM 404 424 POTENTIAL.  
 FT TRANSMEM 427 447 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT TRANSMEM 514 534 POTENTIAL.  
 FT TRANSMEM 543 563 POTENTIAL.  
 FT TRANSMEM 1481 1501 POTENTIAL.  
 FT ACT\_SITE 186 186 POTENTIAL.  
 FT ACT\_SITE 330 330 POTENTIAL.  
 FT SITE 233 233 POTENTIAL.  
 FT SITE 235 235 SUBSTRATE BINDING (POTENTIAL).  
 SQ SEQUENCE 1518 AA; 166404 MW; 0EC99B35B6DE4543 CRC64;

Query Match 4.4%; Score 247.5; DB 1; Length 1518;  
 Best Local Similarity 19.7%; Pred. No. 1.4e-09;  
 Matches 152; Conservative 99; Mismatches 233; Indels 289; Gaps 30;

Db 133 ----- EHELPEDMAQMPVDVFEVSYN---BELSLVSTVIGALDLMPADRLNV 181  
 Qy 375 YVFDGGAAMLSPESLAETSEFAKVVFPCKKYSIEPRADWEYFAAKIDYLDKDVQTSFVK 434  
 Db 182 YIIIDG- 187  
 Qy 435 DRRAMKREYEERIRINALVSKALCPDEGWQDTPWPGNNTGDHGMIOYFLQONGC 494  
 Db 188 -----RRKAFHDFAVEAGA-----GYIIR----- 206  
 Qy 495 LMAEGNELRLVYVSEKRRPGROHKKAGAMALVYSAVLNTPPIILMDCHYINNSK 554  
 Db 207 --AENN-----HAKANLNMLAV---FDSPAVLFDQDQ--VPTRG 241  
 Qy 555 ALREACFLM-DENLKGVCYVOEPORFDGIDKNDRYANRNVYFPIINIRGLGIGOPY 613  
 Db 242 FLRTTGMMADPNL---ALLQTPHIFAPD-----FQRLNLAG--GMIVP--- 282  
 Qy 614 VGTGCVFNRTALYGPPIKVKHKRPSLLSKLGGSRKKSKAKKESDKSRHTDSTV 673  
 Db 283 -PEGNMF-----YGL----- 291  
 Qy 674 PVFNLDIDEGVGAFFDEKALMSQSLERKFGQSAVFVASTLMENGVPSPATPML 733  
 Db 292 -----VODGND-----FMDATFCCSCAI-----I 311  
 Qy 734 LKEAIVHISGVDKSDMGMEIGVSYEDILTFGKMHARGMRISYICMPKLPARKSA 793  
 Db 312 RRVAVVGIG-----GFATETEDAHNLTKMQRGWGVAIYLRPLAA--GLA 356  
 Qy 794 PINLSRLNQLVRLWALGSVEILFSRHCPVIWYGNGLKLEPRAVNTTIYPTTSIPLM 853  
 Db 357 TERLLIHIGQVRWARGMIQIMRLDPMIGAG---LRWEOIRCYLSAMSHLFAIPRLT 412  
 Qy 854 YCTLLAVCLFTNPFITPQSNIASIMEL-SLPLSTFATGILLEKRWGSGVIDEMNRDEQF 912  
 Db 413 FLVSPPLAFLEGNTITTAASPLAISYALPHIEFHSVTLSRTIBRW-----RYSF 462  
 Qy 913 -VIGVSAHLFAVFOGILKVLGIDITNFTKASDSDQDFAEIYLFKWTTLIPPTLL 971  
 Db 463 SEIYERSIALFLIRITIVILLQPHKKFNVT---DKGGLLAGVY-FQMDA--YRPVILA 516  
 Qy 972 IYNLVGVAVGSAINSGYQSGLFLGKLPFAFWIVHLYPLKGL-MGRNR 1023  
 Db 517 GYVCAALLKGVY-FGIWQFHDRLALOSFTLIMLVVISIIVLASIYVGRER 568  
 RESULT 8  
 BCAS\_ACEXY STANDARD; PRT; 1518 AA.  
 AC 09MX75;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative cellulose synthase 3 (includes: Cellulose synthase  
 DE catalytic subunit (UDP-forming) (EC 2.4.1.12); Cyclic di-GMP binding  
 DE domain (Cellulose synthase 3 regulatory subunit)).  
 GN BCSABII-B.  
 OS Acetobacter xylinus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;  
 OC Gluconacetobacter.  
 OX NCBI\_TaxID=28448;  
 RN [1]  
 RP .SEQUENCE FROM N.A.  
 RC STRAIN-JCM 7664 / IFO 13693;  
 RX PubMed=10382968;  
 RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,  
 RA Inoue Y.,  
 RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM  
 RT 7664: implication of a novel set of cellulose synthase genes";  
 RL DNA Res. 6:109-115(1999).  
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))(N) = UDP  
 CC + ((1,4-beta-D-glucosyl))(N+1).



QY	495	LDAGNELRLVYVYSREKRFQFHNKKAGANNALVRSAYVLNCPFLINLDCCHYINNSK	555
Db	207	--AENN-----HAKAGNLNHALAV---TDSPPAVIFDCDH-VLPFG	241
QY	555	ALREAMCFLM-DPMLGKQVCYVDFPQRFQEDIDKNDRYANRNTVFEDINLRGLDIOGPVY	613
Db	242	FLRRITQMMMDRPL-----ALLQRRPHFYARDP-----FQNRNLG--GMHVP--	282
QY	614	VGTCGVNRRLALTYIEPPIKVKKKPSLLSKLGGSKKSKAKKESDKKSGNHDSTVY	673
Db	283	-PEGNMF-----YGL-----	291
QY	674	PVFMULDIEEGVEGAFGDEKALLMSQMSLEKRGOSAVFYASTLMENGVPSPATPENL	733
Db	292	-----VQDND-----FWDAITFCGSCAT-----I	311
QY	734	LKEAIIHVSCEGEDKSDMGMEIGWLYGSVTEIDLTFGRMHARGRSIYCMKPLPAFGSA	793
Db	312	RREAVVMIG-----GRATEVTEADATALKMQRGSGTAYLRPELAA--GLA	356
QY	794	PINISDRNLQVLRKALSSVELLSFRHCPIMYIGYGRKLFERLYVNTTIPITISPLM	853
Db	357	TERLILHIGQVRWRARQMIOIMRIDNPLAG---LRWEQRLCYLSAMSHFLPAIRLT	412
QY	854	YCTLLAVCLNFOPIITPOISNIASIMFL-SLFLSIFPATGILEMRMSGVIGIDEMRNQFV	912
Db	413	PLVSPFLAFLFGQNIILASPLAISVYALPHIFHSVITLRSRGW-----RISFW	462
QY	913	-VIGVSANHLFAVEQGIKVLGAGIDTFYTVSKASDEGQFAELYLRKMTTLIPPTLL	971
Db	463	SEIYETSLAFLPLVRIITVTLTLQPHKGFENVT---DKGGLLARGY-FDWDA-VYPNVILA	516
QY	972	IVNLVGVVAGSVYINSGYOSMGLFQKLPFAFVYIHLIYFLKGL-MGRNR	1023
Db	517	GVLCMAALLRGV-FGIWQFHDRLALQSFILNTLWVLSLIIVIASIVGRETR	568

RESULT\_9

BCA2\_ACCEXY

BCA2\_ACCEXY

STANDARD;

PRT;

756 AA.

AC

082859;

15-JUN-2002 (Rel. 41, Created)

DT

15-JUN-2002 (Rel. 41, Last sequence update)

DT

15-JUN-2002 (Rel. 41, Last annotation update)

DE

Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).

GN

BCSA.

OS

Acetobacter xylinus.

OC

Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;

CC

Glucanacetobacter.

OX

NCBI\_TaxID=28448;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN-BPR 2001;

RX

MEDLINE=98296257; Pubmed=9630539;

RA

Nakai T., Moriya A., Tonouchi N., Tsuchida T., Yoshinaga F.,

RA

Horiouchi S., Sone Y., Mori H., Sakai F., Hayashi T.;

RT

"Control of expression. by the cellulose synthase (bcsA) promoter

RT

region from Acetobacter xylinum BPR 2001.";

RL

Gene F13:93-100(1998).

CC

-1 FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes

CC

uridine 5'-diphosphate glucose to cellulose. The thick cellulosic

CC

matrix generated by this enzyme probably provide a specialized

CC

protective environment to the bacterium (by similarity).

CC

-1 CATALYTIC ACTIVITY: UDP-glucose + ((1,4)-beta-D-glucosyl)(N) = UDP

CC

+ ((1,4)-beta-D-glucosyl)(N+1).

CC

-1 COFACTOR: Magnesium (by similarity).

CC

-1 ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid

CC

(c-di-GMP).

CC

-1 PATHWAY: Bacterial cellulose biosynthesis.

CC

-1 SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane

CC

(potential).

CC

-1 DOMAIN: There are two conserved domains in the globular part of

CC the catalytic subunit: the N-terminal domain (domain A) contains  
CC the conserved DXD motif and is possibly involved in catalysis and  
CC substrate binding. The C-terminal domain (domain B) contains the  
CC QXXW motif and is present only in processive glycosyl  
CC transferases. It could be involved in the processive function of  
CC the enzyme, possibly required for holding the growing glycan chain  
CC in the active site.  
CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: AB010645; BAA31463.1; -  
DR InterPro: IPR003919; CellSynth\_A.  
DR Pfam: PF00535; Glycos\_transf\_2; 1.  
DR PRINTS: PR01439; CELLSTHASEA.  
KW Cellulose biosynthesis; Transferase; Glycosyltransferase;  
KM Transmembrane, inner membrane.  
FT DOMAIN 147 242 CATALYTIC SUBDOMAIN A.  
FT TRANSMEM 319 379 CATALYTIC SUBDOMAIN B.  
FT TRANSMEM 27 47 POTENTIAL.  
FT TRANSMEM 49 69 POTENTIAL.  
FT TRANSMEM 106 126 POTENTIAL.  
FT TRANSMEM 167 187 POTENTIAL.  
FT TRANSMEM 409 429 POTENTIAL.  
FT TRANSMEM 432 452 POTENTIAL.  
FT TRANSMEM 470 490 POTENTIAL.  
FT TRANSMEM 517 537 POTENTIAL.  
FT TRANSMEM 551 571 POTENTIAL.  
FT ACT\_SITE 189 189 POTENTIAL.  
FT ACT\_SITE 335 335 POTENTIAL.  
FT SITE 238 238 SUBSTRATE BINDING (POTENTIAL).  
FT SITE 240 240 SUBSTRATE BINDING (POTENTIAL).  
SQ SEQUENCE 756 AA; 84561 MW; 6954F39A25E73B0A CRC64;  
  
Query Match 4.28; Score 236.5; DR 1; Length 756;  
Best Local Similarity 19.38; Pred. No. 3; Le-09;  
Matches 166; Conservative 98; Mismatches 263; Indels 331; Gaps 34;  
  
204 EKNTGCVSTQASERGVDAIDSTD---ILADEALLNDEAR-----OPL 244  
11 ESRIGTISNKKILSRASTYVGLGICALIATTVTLNNEQLVAACVVFVYGRGK 70  
245 SRRVSIPISSRINPYRMVIMLRLVILCLFLHYRTNPVNAFALML-----SVIGEIMF 298  
71 SRRTOI-----FLEVLALVSLRYLWRLTEITLD-FNWIGIGILVILMAELVA 119  
299 ALSMIIDOPPKPVPVRETYLDRLARYREGESPOLAVIDIVSVDPKEPPLVTANT 358  
120 LYLWLFSTYQTIQPLHRAIP---LPLPDVNDWMP---VDIPIPYD---EQLSIVRLT 168  
359 VLSILAVDPVDKVCYFDDGAAMLSESLAETSEFARKWVFCCKYSIEPRAPMYFA 418  
169 VIGALGIDMPDKVNVYIIDDG----- 190  
419 AKIDYLDKVVQVSEFKDRRAKREVEFKIRINALVSKLCPCEGWMQDGTWPGNNT 478  
191 -----VRPEFQF-----AKDC----- 202  
479 GDHPGMIOYFLONGGLDAENGLRPLVYVSRKRGFGHKKAGAMALVRSAYLVNG 538  
203 ---GALYI---GAVDVA-----SAHAAAGMLNNAIK---RTSG 231  
539 PFLIADCDHYIINNSKALREAMFLMDPNLGKQVQYQPPQRFQIDKKDRYANRTVEF 598  
232 DYILLIDCDH-IPTRAFLOIAMGMVY---ADRKIALMQPPIPHHYSPDP----- 275

QY 599 DINKRLDGIQGFVYVGTGCVENRTALVYEPIKVKHKKPSILSKLGGSRKKSKAKK 658  
DB 276 -----FQRLAVGYRTF----- 287  
QY 659 ESDKKSGRHTDSFVFNPDIEEGVEGAFDEKALLMSQMSLEKRFQSGAVFVASTL 718  
DB 288 -----PEGNL---FYGVIODGN-----FMDATFFGSCA 314  
QY 719 MGVGPSATPENLKEAIIHVISCYEDKSDMGMEIGNYGVSTEDILTGFMARGMR 778  
DB 315 I-----LRRALIESIG-----GFAVEVETDATHALMQRGMS 348  
QY 779 SIYCKMLPAFKSGAPINSDRLNOLYRLMAGSVELFESRHPCTWGYGR-LKFLERRA 837  
DB 349 TAYL-RIPVASGLATERLTTHGQHMRAWMIOI-FRDNMML---GRGILKQRLC 401  
QY 838 YVNTIYPIITSIPLWYCLLWCLTQNFIIPOINMSIMWLSFLSIFANGILEMR 897  
DB 402 YLSAMTSFPPALPRVIFLAPLAFRAGONTIAADPLAVALPMPHSHAL----- 454  
QY 898 SGVIDEMWNRNEQFW-VIGGSANLFAVFGILKVLADITNTVTYSKA---SDEGDPA 953  
DB 455 -AAKVKKGR-YEFNSEVYETTMALPLVAVTITLFLPCKKFNTEKGGVLEEFEDIG 512  
QY 954 ELY-----LEKWTLLIPTTILLI-VNLGVAVGVYVAINGYQSGPLPKLPAPV 1006  
DB 513 ATVPNIIFATIMWGILIGFELIVAFNOLDVYARNALYNCA-----WA 557  
QY 1007 IYHLVFLKGL-MGRQNR 1023  
DB 558 LISLITLFAIAVARETK 575  
  
RESULT 10  
BCAL\_ACEXY STANDARD; PRT; 754 AA.  
ID BCAL\_ACEXY  
AC P19449;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).  
GN BCSA.  
OS Acetobacter xylinus.  
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;  
OC Gluconacetobacter.  
OX NCBI\_TaxID=28448;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-42 AND 196-206.  
RC STRAIN-1306-3;  
RX MEDLINE-91045951; Pubmed-2146681;  
RX Wong H.C., Fear A.L., Calhoun R.D., Eichinger G.H., Mayer R.A.  
RA Amikem D., Benzinan M., Gelfand D.H., Meade J.H., Emerick A.W.,  
RA Bruner R., Ben-Bassat A., Tal R.;  
RT "Genetic organization of the cellulose synthase operon in Acetobacter  
RT xylinum";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:8130-8134(1990).  
RN [2]  
RP ENZYME REGULATION.  
RC STRAIN-1306-3;  
RX Pubmed-11297407;  
RX Chang A.L., Tuckerman J.R., Gonzalez G., Mayer R., Weinhouse H.,  
RA Volman G., Amikem D., Benzinan M., Gilles-Gonzalez M.A.;  
RT "Phosphodiesterase A1, a regulator of cellulose synthesis in  
RT Acetobacter xylinum, is a heme-based sensor";  
RL Biochemistry 40:3420-3426(2001).  
CC - FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes  
CC uridine 5'-diphosphate glucose to cellulose in a processive way.  
CC The thick cellulosic mats generated by this enzyme probably  
CC provide a specialized protective environment to the bacterium.  
CC - CATALYTIC ACTIVITY: UDP-glucose + ((1,4)-beta-D-glucosyl)(N) = UDP  
CC + ((1,4)-beta-D-glucosyl)(N+1).  
CC - COFACTOR: Magnesium.  
CC - ENZYME REGULATION: Activated by bis-(3'-5') cyclic diuanylic acid

CC (C-di-GMP).  
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (potential).  
 CC -1- DOMAIN: There are two conserved domains in the globular part of  
 CC the catalytic subunit: the N-terminal domain (domain A) contains  
 CC the conserved DXD motif and is possibly involved in catalysis and  
 CC substrate binding. The C-terminal domain (domain B) contains the  
 CC OXXRW motif and is present only in processive glycosyl  
 CC transferases. It could be involved in the processivity function of  
 CC the enzyme, possibly required for holding the growing glycan chain  
 CC in the active site.  
 CC -1- INDUCTION: Cellulose is produced at a linear rate with respect to  
 CC cell growth when O(2) is present.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 CC EMBL: M37202; AAA21884.1; -  
 CC PIR: A43735; A43735.  
 CC InterPro: IPR003919; Cellsynth\_A.  
 CC InterPro: IPR001173; Glycos\_transf\_2.  
 CC Pfam: PF00535; Glycos\_transf\_2; 1.  
 CC PRINTS: PR01439; CELLSTHASEA.  
 CC Cellulose biosynthesis; Transferase; Glycosyltransferase;  
 CC Transmembrane; Inner membrane.  
 CC FT DOMAIN 147 240 CATALYTIC SUBDOMAIN A.  
 CC FT 317 377 CATALYTIC SUBDOMAIN B..  
 CC FT TRANSMEM 26 46 POTENTIAL.  
 CC FT TRANSMEM 47 67 POTENTIAL.  
 CC FT TRANSMEM 108 128 POTENTIAL.  
 CC FT TRANSMEM 167 187 POTENTIAL.  
 CC FT TRANSMEM 407 427 POTENTIAL.  
 CC FT TRANSMEM 430 450 POTENTIAL.  
 CC FT TRANSMEM 468 488 POTENTIAL.  
 CC FT TRANSMEM 516 536 POTENTIAL.  
 CC FT TRANSMEM 549 569 POTENTIAL.  
 CC FT ACT\_SITE 189 189 POTENTIAL.  
 CC FT ACT\_SITE 333 333 POTENTIAL.  
 CC FT SITE 236 236 SUBSTRATE BINDING (POTENTIAL).  
 CC FT SITE 238 238 SUBSTRATE BINDING (POTENTIAL).  
 CC SO SEQUENCE 754 AA; 84442 MW; 8D5FC1FE62E2C068 CRC64;  
 CC -----  
 CC Query Match 4.0%; Score 228.5; DB 1; Length 754;  
 CC Best Local Similarity 19.3%; Pred. No. 1.2e-08;  
 CC Matches 154; Conservative 84; Mismatches 229; Indels 329; Gaps 30;  
 CC -----  
 CC Oy 263 MLRLVILCLFLHYRTNPVPNAFMLT-----SVIGEIFALSMILDOFPKPPVRE 316  
 CC Db VLSALVSLRYLTWRLETFILD--FDTWIOGGGLVTLMAELVALMLFISYFTQIPLRHA 137  
 CC Oy 317 TYDLRLALRYDREGPSQLAVDIFVSTVDPLKEPPLVTANTVLSILAVDVPVKVSCYV 376  
 CC Db 138 P-----LPLPDVNDVDP-----VDLFIPTYD--EQLSTVRLVIGALGIDMPDPVNYVI 186  
 CC Oy 377 FDDGAAMLFSFSLAETSEFAKKWVPFCCKYSIEPRAPMYFAAKIDYLKDKVQTSFVDR 436  
 CC Db 187 LDDG-----||| 190  
 CC Oy 437 RAMKREYEEFKIRINALVSKALKCPDEGVMQDGTMPDGNNTGHPGMIOVFLGONGILD 496  
 CC Db 191 --VPEFEQFAKDCAL-----||| 205  
 CC Oy 497 AEGNLEPLVYVSRKRGFOHKKAGAMALVRYSAVLTNCPFLTLNDCDHYTNNSKAL 556  
 CC Db 206 -----YIGRVD-----SSHAKAGMLNHAIK-----RTSGDYTLILDCDH-IPTRAF 246

Oy 557 REAMCELMADPNLKGVCYVQFPQFDGIDKNDRYANRMTVEFDINLRGLDGIQGVYVGT 616  
 Db QIAMQMWV---ADRIALMQLPHHYSPDP----- 273  
 Oy 617 GCVFNRTALYGEPIKYYKKKPSLLSKLGGSSRRKNSKAKKESDKSGRHTDSTVPVF 676  
 Db 274 ---FQRLNLAVGRTD----- 288  
 Oy 677 NDDIDEEGVGAGFDEKALLMSQMSLEKRFQSAVFAVASTLMENGVPSPATPENLKE 736  
 Db 289 NL---FYGIQDGN-----FMDATFFGSCAI-----LRE 317  
 Oy 737 AIHVISCGEDKSDGMETIGVSGVTEIDILGFHARGWSIYCMPLPFAKGSAPIN 796  
 Db 318 AIESIG-----GFAVEYTEDAHIALMKRQGWSTAIL--RIPASGLATPR 362  
 Oy 797 ISDRINQVLRMALGSEVILFSRHCPHWGYNRLFLERFAVNTTYPITSIPLMVCT 856  
 Db 363 LTHHGQRMRMRMARGIQI-FRVDNPM---LGGLKLQGRCLYLSAMTSFFAIPRVIFLA 418  
 Oy 857 LLAVLFTNQFIIPDISNASTWFLSLFISFATGILEKRWMS---CVGIDEMWRNDEQW 912  
 Db 419 SPLAEFLFGQNLIIAASP-----LAVLAVALPHMFHSIATRAKYNKQWR-YSEWS 466  
 Oy 913 VIGYSANLFAVFGCILKVLGIDINFTVTSKA---SDRDGFALTYLFKMTTLIPPT 969  
 Db 467 EYETTMALFLVAVITITLMPFSKGFNTGCVLEEFELGATY-----PI 516  
 Oy 970 LLIIVNLGVVAGVSAINSYGSGWPLFG-KLFFAVVYVHLYPELKGMRONTPTI 1027  
 Db 517 IF-----AGIMTGLIGLGFELTFHF-----NQLAGIAKRAYLL 550  
 Oy 1028 VVWVS-----VLLASI 1038  
 Db 551 NCIMAMISLIIILAAI 566  
 RESULT 11  
 BCSA\_PSEFL STANDARD; PRT; 739 AA.  
 AC P58931;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cellulose synthase catalytic subunit [UDP-forming] (EC 2.4.1.12).  
 GN BCSA OR WSSB.  
 OS Pseudomonas fluorescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=294;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SBW25;  
 RX MEDLINE=22013850; PubMed=12019221;  
 RA Spliers A.J., Kahn S.G., Bohannon J., Travisano M., Rainey P.B.;  
 RT "Adaptive divergence in experimental populations of Pseudomonas  
 RT fluorescens. I. Genetic and phenotypic bases of wrinkly spreader  
 RT fitness.";  
 RL Genetics 161:33-46(2002).  
 CC -1- FUNCTION: Catalytic subunit of cellulose synthase. It polymerizes  
 CC uridine 5'-diphosphate glucose to cellulose, which is produced as  
 CC an extracellular component responsible for the structural  
 CC integrity and rigidity of self-supporting mats characteristic of  
 CC the "wrinkly spreader" phenotype.  
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + ((1,4-beta-D-glucosyl))<sub>(N)</sub> = UDP  
 CC + ((1,4-beta-D-glucosyl))<sub>(N+1)</sub>.  
 CC -1- COFACTOR: Magnesium (By similarity).  
 CC -1- ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid  
 CC (C-di-GMP) (By similarity).  
 CC -1- PATHWAY: Bacterial cellulose biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -1- DOMAIN: There are two conserved domains in the globular part of

CC the protein: the N-terminal domain (domain A) contains  
 CC the conserved DXD motif and is possibly involved in catalysis and  
 CC substrate binding. The C-terminal domain (domain B) contains the  
 CC QXXRW motif and is present only in processive glycosyl  
 CC transferases. It could be involved in the processivity function of  
 CC the enzyme, possibly required for holding the growing glycan chain  
 CC in the active site.  
 CC -1 SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: AY074776; AAL71842.1; -  
 KW Cellulose biosynthesis; Transferase; Glycosyltransferase;  
 KM Transmembrane; Inner membrane.  
 FT DOMAIN 157 250  
 FT TRANSMEM 327 387 CATALYTIC SUBDOMAIN A.  
 FT TRANSMEM 36 55 CATALYTIC SUBDOMAIN B.  
 FT TRANSMEM 59 76 POTENTIAL.  
 FT TRANSMEM 83 101 POTENTIAL.  
 FT TRANSMEM 116 138 POTENTIAL.  
 FT TRANSMEM 417 436 POTENTIAL.  
 FT TRANSMEM 440 462 POTENTIAL.  
 FT TRANSMEM 524 546 POTENTIAL.  
 FT TRANSMEM 551 573 POTENTIAL.  
 FT ACT\_SITE 199 199 POTENTIAL.  
 FT ACT\_SITE 343 343 POTENTIAL.  
 FT SITE 246 246 SUBSTRATE BINDING (POTENTIAL).  
 FT SITE 248 248 SUBSTRATE BINDING (POTENTIAL).  
 SQ SEQUENCE 739 AA; 82165 MW; 28962EA3854B2BB CRC64;  
 Query Match 4.0%; Score 228; DB 1; Length 739;  
 Best Local Similarity 18.8%; Pred. No. 1.3e-08;  
 Matches 149; Conservative 87; Mismatches 213; Indels 342; Gaps 33;  
 257 PYRWYIMLRVLTIC-----FLHYRTNPVPAALML-----VSICELWAL----- 300  
 DB 80 PERLAI-LALVLSLVSALRIFWRLITSL--GFEIWDMEFGYGLVAEFYALVLTIG 136  
 QY 301 ---SWILDQFKMPEPVNRETYLRLATRYDREGPSQLAADVSYVDPIKEPPLVTA 356  
 DB 137 YVQTAMPLRRTPVWL-----KTEPEEMPTVDVFIPTYN--EALSIYK 176  
 QY 357 NTVLSILANDYVDKVSQYVEDGAAMLSFESLAEISFARKWVPCKKYSIEPPAPEWY 416  
 DB 177 LTFIAQAQMDWPKDKLRVHVLDDG-----RRDDPRE--FCRKVGYN----- 215  
 QY 417 FAAKIDYLDKVQTSFVKDRBRAMKREYEEFKIRINALVSKALCPDEGVMODGTPWPGN 476  
 DB 216 ----- 215  
 QY 477 NTGDHPGMIOVFLGONGGLDAEGNELPRLVYVSREKRPGEQHHKAGANALVRYSAYLT 536  
 DB 216 -----YIRRDV--NPHAKAGNLEALKV---T 237  
 QY 537 NCPFLINDCDHYINNSKALREAMGFLMDPNLKGQVQYQFQRF--DGIKN-DRY-- 590  
 DB 238 DEEYIALDPADHVPTRSLQVSLGFWLKDPKL---AMQTPHFEPSPDPFKNDLTPRA 293  
 QY 591 -ANRNTVFEDINLRGLDGIQGVYVGTGVFNRTALYGEPLIKVHKRPILLSLKGGS 649  
 DB 294 VNEEELFGLVQDNDLMAATFFGSCAVIR-----EP----- 328  
 QY 650 RKNRSKAKESDKKSGRHTDSTVPVFNDDIEEGVEGAFDEKALIMSQMSLEKRFQ 709  
 DB 329 ----- 328  
 QY 710 SAVFVASTLMENGVPSPATPENLKAHVISCYEDKSDMGMEIGWYGSVTEIDILTG 769

DB 329 -----LLEIGV-----AVE-----YTTEDAHTA 347  
 QY 770 FPMHANGMSIYCMPLPAFKSAPINLSDDLNOVRLMAGSEILFSRHCPIWGYNR 829  
 DB 348 LKLNRIQVYATL--AIPQAGLATESLRHNRORRMAGAOIFRTDPLLGKC---- 401  
 QY 830 LKLEFAVNTIYPTISPIPLMWCTL-LAVCLEPTNPIIPIQSIASIWFSLEPL-SI 887  
 DB 402 LKMGORICANMGMHFEYLPRLVFLTAPLALVIG-----AEIFASLMIAVAVLPHL 456  
 QY 888 FATGILEMRSGVGIDEMWRNQF-----WVIGVSAHLFAVFOGLKVLADINFTYT 942  
 DB 457 VHSSTLNSRIQGRFRHSFV-NEUYETVLAWYI-----LPYLVALVNPKAG--GFNTY 506  
 QY 943 SKASDEGDFAELIEFKWTLLIPTTLIVLVGVAGVSAVINSQSGMGLGKLPF 1002  
 DB 507 DKGIIDKQF-----FDW-KLARPYLVLAIVNLIGFGIHOLI-----WGDASTAVTV 554  
 QY 1003 AF---WVIYHL 1010  
 DB 555 AINLFWTLXNL 565  
 RESULT 12  
 ID BCAS3\_ACEXY STANDARD; PRT; 745 AA.  
 AC 09WX61;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cellulose synthase 1 catalytic subunit [UDP-forming] (EC 2.4.1.12).  
 GN BCSAI.  
 OS Acetobacter xylinus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;  
 OC Gluconacetobacter.  
 OX NCBI\_TaxID=28448;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JCM 7664 / IFO 13693;  
 RX PubMed=10382968;  
 RA Umeda Y., Hiraio A., Ishibashi M., Akiyama H., Onituka T., Iseuchi M.,  
 RA Inoue Y.;  
 RA "Cloning of cellulose synthase genes from Acetobacter xylinum JCM  
 RT 7664: implication of a novel set of cellulose synthase genes";  
 RL DNA Res. 6:109-115(1999).  
 CC -1 FUNCTION: Catalytic subunit of cellulose synthase. The thick cellulosic  
 CC uridine 5'-diphosphate glucose to cellulose. The thick cellulosic  
 CC mats generated by this enzyme probably provide a specialized  
 CC protective environment to the bacterium (By similarity).  
 CC + ((1,4)-beta-D-glucosyl)(N+1).  
 CC -1 CATALYTIC ACTIVITY: UDP-glucose + ((1,4)-beta-D-glucosyl)  
 CC + ((1,4)-beta-D-glucosyl)(N+1).  
 CC -1 COFACTOR: Magnesium (By similarity).  
 CC -1 ENZYME REGULATION: Activated by bis-(3'-5') cyclic diguanylic acid  
 CC (c-di-GMP) (By similarity).  
 CC -1 PATHWAY: Bacterial cellulose biosynthesis.  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Potential).  
 CC -1 DOMAIN: There are two conserved domains in the globular part of  
 CC the catalytic subunit: the N-terminal domain (domain A) contains  
 CC the conserved DXD motif and is possibly involved in catalysis and  
 CC substrate binding. The C-terminal domain (domain B) contains the  
 CC QXXRW motif and is present only in processive glycosyl  
 CC transferases. It could be involved in the processivity function of  
 CC the enzyme, possibly required for holding the growing glycan chain  
 CC in the active site.  
 CC -1 SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, AB015802; BAA77565.1;
DR  InterPro: IPR003919; CellSynth_A.
DR  InterPro: IPR001173; Glycos_transf_2.
DR  Pfam: PF00535; Glycos_transf.2; 1.
DR  PRINTS: PR01439; CELLSTHASEA.
KW  Cellulose biosynthesis; Transferase; Glycosyltransferase;
FT  Transmembrane, inner membrane.
FT  DOMAIN 147 240
FT  DOMAIN 317 377
FT  TRANSMEM 29 49
FT  TRANSMEM 106 126
FT  TRANSMEM 153 173
FT  TRANSMEM 407 427
FT  TRANSMEM 430 450
FT  TRANSMEM 468 488
FT  TRANSMEM 515 535
FT  TRANSMEM 547 567
FT  TRANSMEM 649 669
FT  ACT_SITE 189 189
FT  ACT_SITE 333 333
FT  SITE 236 236
FT  SITE 238 238
SQ  SEQUENCE 745 AA; 83518 MM; 57EA0457A226F815 CRC64;

Query Match 3.7%; Score 208.5; DB 1; Length 745;
Best Local Similarity 18.9%; Pred. No. 3.3e-07;
Matches 149; Conservative 82; Mismatches 235; Indels 321; Gaps 31;

OY 263 MLRLVILCLFLHYRTNPVNAFLM-----LVSVEIMWALSMIDQPKPVNRE 316
DB 80 VLSALVSLRYLFWRLFTLD--FDTWGILGLVLLLELVALYMLFSYFOTISPLHRA 137
OY 317 TYDLRLALRYDREGESQLAIVDFSTVDPLKEPPLTANTVLSILAVDYVDKVCYV 376
DB 138 P-----LPL-----PANPDEMPYDIFIPYD--ELSLYRLVLGALGIDMPDPKVNVI 186
OY 377 FDDGAAMLSPESLSTSPARKWVPCCKYSTEPAPEMVFAKIDYLDKQVQTSFVKDR 436
DB 187 LDDG-----RREFAR-----FA----- 199
OY 437 RAMKREVEFKIRIALVSKALKCEBGMVMDGTPWPGNNGDHPGMIOVFLQNGGID 496
DB 200 -----E 200
OY 497 AEGNELPRLVYVRSREKRGFQHHKKAGAMALVRSAYLTNGPFIILDCDHYINNSKAL 556
DB 201 ACGAR-----YIARDNA-----HAKAGNLNVAIK-----HTTGDHLLIDCDH-IFTRAF 246
OY 557 REAMCELDPNLKGOCYVQFPQRFQIDKNDRYANRNTVFPDNLRLGLDIGQPVYVGT 616
DB 247 QISMGMWVSDS-----NIALLOTPHNHFSPDP----- 273
OY 617 GCVENRTALGYEPIRYKHKHKKPSLLSKLGGSRKKKSKAKESDKKSGRHTDSVPVF 676
DB 274 ---FORNLAVGTRP----- 288
OY 677 NLDDIEGVEGAGFDEKALMSQMSLEKRFQSAVFVASTLMENGVPSPATPENLKE 736
DB 289 NL---FYGVLDGND-----FWDAFFFGSCAI-----LRKK 317
OY 737 AIHWISGCEYEDKSDGMEIGWYGVTEDDILGTFMHAARGWSIYCMPLPAFKGSAPIN 796
DB 318 AIEEIG-----GFATEVTEDAHTALMOKRGWSTAVL--RIPLASGLATFR 362
OY 797 LSDRLNOVLRNALSGVELLFSRHCPIWYNGYNGRLKFLBFAVYNTTIPIITSIPLMKCT 856
DB 363 LITHIGQRRKMRARGIQI--FRVDNPM---LGGSLKLGQGLCLSLMTSFEFFAIPRYFLA 418
OY 857 LLAVCLFTNOFTIPIQISNIASIMFLSLFISFATGILEMRWMSGVGVDEWMRNEQFW--VI 914

```

---

```

DB 419 SPLAEFLFSQNIITAA SPLAVGYAIPHMFHSIAT-----AAKNKGM-RYSEWSEYV 469
OY 915 GGVSAHLRAVFGGILAKVLADIDNTFTVNSKA---SDEGDFAELVLFKFTTLIPPTLL 971
DB 470 ETVMA-LFLVRTVITVMTLFPSPKKNVEKGVLEKEEFDLATY-----PNIIFA 519
OY 972 IYNLVGVAVG-----SYAINSGQSMGFLGKLFFAFWIVHLYPLKGL 1017
DB 520 IIMALGLGLYALLFQHLIDISERAYALN-----CIWVSILITIMAVI 564
OY 1018 -MGRNR 1023
DB 565 SVGREK 571

RESULT 13
ACSI_ACEXY STANDARD; PRT; 1550 AA.
ID ACSI_ACEXY
AC P21877; P37717;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose synthase 1 [includes: Cellulose synthase catalytic domain
DE [UDP-forming] (EC 2.4.1.12); Cyclic di-GMP binding domain (Cellulose
DE synthase 1 regulatory domain)].
CN ACSAB OR ACSA OR ACSB.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconacetobacter.
OX NCBI_Taxid=28448;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 6-20.
RC STRAIN=ATCC 53582;
RA MEDLINE=91346705; PubMed=2151718;
RX Saxena I.M., Lin F.C., Brown R.M. Jr.;
RT "Cloning and sequencing of the cellulose synthase catalytic subunit
RT gene of Acetobacter xylinum.";
RL plant Mol. Biol. 15:673-683(1990).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 768-781.
RC STRAIN=ATCC 53582;
RA MEDLINE=91322509; PubMed=1830823;
RX Saxena I.M., Lin F.C., Brown R.M. Jr.;
RT "Identification of a new gene in an operon for cellulose biosynthesis
RT in Acetobacter xylinum.";
RL plant Mol. Biol. 16:947-954(1991).
RN [3]
RP REVISIONS.
RC STRAIN=ATCC 53582;
RX PubMed=8083166;
RA Saxena I.M., Kudlicka K., Okuda K., Brown R.M. Jr.;
RT "Characterization of genes in the cellulose-synthesizing operon (acs
RT operon) of Acetobacter xylinum: implications for cellulose
RT crystallization.";
RL J. Bacteriol. 176:5735-5752(1994).
RN [4]
RP SEQUENCE OF 1-8 FROM N.A.
RC STRAIN=ATCC 23769;
RX MEDLINE=94131945; PubMed=8300521;
RA Standaal R., Iversen T.-G., Coucheron D.H., Fjaervik E., Blatny J.M.,
RA Valla S.;
RT "A new gene required for cellulose production and a gene encoding
RT cellulolytic activity in Acetobacter xylinum are colocalized with the
RT bcs operon.";
RL J. Bacteriol. 176:665-672(1994).
RN [5]
RP FUNCTION.
RC STRAIN=ATCC 53582;
RX PubMed=2138620;
RA Lin F.C., Brown R.M. Jr., Drake R.R. Jr., Haley B.E.;
RT "Identification of the uridine 5'-diphosphoglucose (UDP-glc) binding
RT subunit of cellulose synthase in Acetobacter xylinum using the
RT photoaffinity probe 5-azido-UDP-glc.";

```



QY 958 FKWTTLLIPTLLIVNLGVAVSYAINSGYSGWPLFGKLFEPFVIVHLYPFLKGL 1017  
 Db 510 -----GAVYPIIILGLIFGGLARGV-YELSGHLDQIAERAYLLNSAMAMSLIIILAI 564  
 QY 1018 -MGRO-----NRPPIVVV 1030  
 Db 565 AVGRETOOKRNSHRIPIATIPV 585

RESULT 14  
 COX1\_PARLI STANDARD; PRT; 517 AA.  
 ID COX1\_PARLI STANDARD; PRT; 517 AA.  
 AC P12700;  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).  
 GN COI.  
 OS Paracentrotus lividus (Common sea urchin).  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Echinidae;  
 OC Paracentrotus.  
 OX NCBI\_TaxID=7656;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-69291831; PubMed-2544576;  
 RA Cantatore P., Roberti M., Rainaldi G., Gadaleta M.N., Saccone C.;  
 RT "The complete nucleotide sequence, gene organization, and genetic  
 code of the mitochondrial genome of Paracentrotus lividus.";  
 RL J. Biol. Chem. 264:10965-10975(1989).  
 RN [2]  
 RP SEQUENCE OF 469-517 FROM N.A.  
 RC TISSUE-Egg;  
 RX MEDLINE-87248108; PubMed-3596250;  
 RA Cantatore P., Roberti M., Morisco P., Rainaldi G., Gadaleta M.N.,  
 RA Saccone C.;  
 RT "A novel gene order in the Paracentrotus lividus mitochondrial  
 genome.";  
 RL Gene 53:41-54(1987).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
 AND COPPER B.  
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome  
 c + 2 H(2)O.  
 CC -1- PATHWAY: Respiratory chain; terminal step.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J04815, AAA68135.1, -;  
 DR EMBL: M16519, AAA31989.2, -;  
 DR PIR: C34284, C34284.  
 DR HSSP: P00396, 20CC.  
 DR InterPro: IPR000883, COX1.  
 DR Pfam: PF00115, COX1.1.  
 DR PRINTS: PR01165, CYCOXIDASEI.  
 DR PROSITE: PS00077, COX1.1.  
 DR Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
 KW Respiratory chain; Inner membrane.

FT METAL 61 61 IRON (HEME A) (PROBABLE).  
 FT METAL 240 240 COPPER B (PROBABLE).  
 FT METAL 244 244 COPPER B (PROBABLE).  
 FT METAL 290 290 COPPER B (PROBABLE).  
 FT METAL 291 291 COPPER B (PROBABLE).  
 FT METAL 376 376 IRON (HEME A3) (PROBABLE).  
 FT METAL 378 378 IRON (HEME A) (PROBABLE).  
 SQ SEQUENCE 517 AA; 57217 MW; BA2A1P1D5789A856 CnC64;

Query Match 2.1%; Score 119; DB 1; Length 517;  
 Best Local Similarity 20.3%; Pred. No. 0.63;  
 Matches 87; Conservative 56; Mismatches 139; Indels 146; Gaps 21.

QY 724 VPPSATPENLKEAHIVHSICGEDSKDGMELGW-ITYGSYEDI-----LNGFKNA 774  
 Db 105 IPPSF-----ILLASAGVES---GAGTGWITPPLSNNAHAGSDVLAIFSJHL 152  
 QY 775 RGMRSIYC-----MPKLPFGSAPINLSD 799  
 Db 153 AGASSILASINFTTIIMRTPGMSFDRLPLFVMSVFYATFLLLSLPVLGATIMLLTD 212  
 QY 800 R-----LNQVLRNALGSVEL-----FSRHCPITWYGNRLAFLERFA 837  
 Db 213 RNINTTFEDPAGGDPILFQHLFWFEGHPEYVILLPFGMISHVIAHYSK---REDFG 269  
 QY 838 YVNTTIYPTISPIPLMCTLLAVCLFTNQFIIPQSNASISWFLSFLISFATGILEMRW 897  
 Db 270 YLG-MVYAMIAIGVGF-LVNAHMF---GMDVDTRAYTAAMITIAV-PTGIKPSW 323  
 QY 898 SGV--GIDEMRNREQFVIGVSANLFAVFOGILKVLGIDPNTFTVTSKASDEGDAEL 955  
 Db 324 MATIGSNIQWETPLMALGFVPLETGLGIVLANSID----- 364  
 QY 956 YLFKWTTLLIPTLLIVNLGVAV-GSVAINSGYSGWPLF-----GKL-PPAF 1004  
 Db 365 -----VLHDYTVVAHFHYVLSMGAVFAIFAGFTMPFCGYNLPLMGKHPFM 417  
 QY 1005 WVIHLYPF-----LKLIMRONRPTPIVVMSV-----LLASIF--SLWVRID 1047  
 Db 418 FVGYNLTFPPQHFGLAGMPRYSYDPATVIMTVSSIGSTISLVNMLFFFLIM---E 474

QY 1048 PFSRWVG 1055  
 Db 475 AFASQREG 482

RESULT 15  
 YKTA\_CAEEL STANDARD; PRT; 590 AA.  
 ID YKTA\_CAEEL STANDARD; PRT; 590 AA.  
 AC P34322;  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C07A9.11 in chromosome III.  
 GN C07A9.11.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Peloderidae; Caenorhabditis.  
 CC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Beirs M.,  
 RA Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kerhaw J., Kirsten J., Laisster N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
 RA Sims M., Smailson N., Smith A., Smith M., Sonnenhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Watson L., Wilkinson-Sproat J.,





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 16, 2003, 11:04:03 ; Search time 49 Seconds  
(without alignments)  
4478.371 Million cell updates/sec

Title: AAC39336  
Perfect score: 5677  
Sequence: 1 MSEGGETAGKPMKNIVPQTC.....IDPFTSRVTPDILGEGINC 1065

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*  
1: sp.\_archaea:\*  
2: sp.\_bacteria:\*  
3: sp.\_fungi:\*  
4: sp.\_human:\*  
5: sp.\_invertebrate:\*  
6: sp.\_mammal:\*  
7: sp.\_mhc:\*  
8: sp.\_organelle:\*  
9: sp.\_phage:\*  
10: sp.\_plant:\*  
11: sp.\_rodent:\*  
12: sp.\_virus:\*  
13: sp.\_vertebrate:\*  
14: sp.\_unclassified:\*  
15: sp.\_viral:\*  
16: sp.\_bacteriophage:\*  
17: sp.\_archaeophage:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5677	100.0	1065	10	048948	048948 arabidopsis
2	5654	99.6	1065	10	09FHK6	09FHK6 arabidopsis
3	4954	87.3	1067	10	09XGX6	09XGX6 gossypium h
4	4571	80.5	1077	10	09LLI6	09LLI6 zea mays (m
5	4564	80.4	1079	10	09LLI1	09LLI1 zea mays (m
6	4529.5	79.8	1076	10	09LLI5	09LLI5 zea mays (m
7	4028.5	71.0	1075	10	09LLI9	09LLI9 zea mays (m
8	4024	70.2	1074	10	09LLI8	09LLI8 zea mays (m
9	3984	70.2	1081	10	048946	048946 arabidopsis
10	3900	68.3	1059	10	09LLI4	09LLI4 zea mays (m
11	3875.5	68.3	1066	10	09LLI3	09LLI3 zea mays (m
12	3843.5	67.7	1094	10	09LLI2	09LLI2 zea mays (m
13	3811	67.1	1026	10	09SWW6	09SWW6 arabidopsis
14	3805	67.0	1026	10	09XHP6	09XHP6 arabidopsis
15	3801	67.0	1065	10	09SKJ5	09SKJ5 arabidopsis
16	3741	65.9	1043	10	09FNC3	09FNC3 arabidopsis

17	3719	65.5	1091	10	093XQ1	093XQ1 nicotiana a
18	3700	65.2	1063	10	09AV71	09AV71 oryza sativ
19	3670	64.6	1069	10	09FTB9	09FTB9 arabidopsis
20	3663	64.5	1042	10	08I649	08I649 populus x c
21	3647.5	64.3	1084	10	09FGF9	09FGF9 arabidopsis
22	3628	63.9	1084	10	048947	048947 arabidopsis
23	3612.5	63.6	1081	10	065338	065338 arabidopsis
24	3590	63.2	1088	10	09SJ22	09SJ22 arabidopsis
25	3522	62.0	821	10	09LLI7	09LLI7 zea mays (m
26	3473.5	61.2	974	10	P93155	P93155 gossypium h
27	3449.5	60.8	978	10	09AXK0	09AXK0 zinnia eleg
28	3406.5	60.0	974	10	08W1M0	08W1M0 gossypium h
29	3404	60.0	985	10	09C5Z8	09C5Z8 arabidopsis
30	3327	58.6	978	10	081368	081368 populus tre
31	3317.5	58.4	958	10	09SN37	09SN37 arabidopsis
32	3114	54.9	939	10	0943H3	0943H3 oryza sativ
33	2808	49.5	685	10	P93156	P93156 gossypium h
34	2325.5	41.0	1145	10	09M9M4	09M9M4 arabidopsis
35	2324	40.9	1181	10	09SRW9	09SRW9 arabidopsis
36	2303.5	40.6	1145	10	09LFL0	09LFL0 arabidopsis
37	2302	40.5	1170	10	09LH27	09LH27 oryza sativ
38	2289.5	40.3	1111	10	09S2L9	09S2L9 arabidopsis
39	2255.5	39.7	1127	10	08W3F9	08W3F9 oryza sativ
40	2246.5	39.6	1127	10	093XQ0	093XQ0 nicotiana a
41	2221.5	39.1	1036	10	049323	049323 arabidopsis
42	2185.5	38.5	979	10	09FVR3	09FVR3 arabidopsis
43	1931	34.0	583	10	048961	048961 oryza sativ
44	1874.5	33.0	507	10	093YF8	093YF8 arabidopsis
45	1871.5	33.0	504	10	09AXJ9	09AXJ9 zinnia eleg

## ALIGNMENTS

## RESULT 1

ID	048948	PRELIMINARY;	PRT;	1065 AA.
AC	048948:			
DT	01-JUN-1998	(TREMBLrel. 06, Created)		
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	Cellulose synthase catalytic subunit.			
GN	AtCB.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RY	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RC	MEDLINE=9811412; PubMed=9445479;			
RA	Artoli T., Peng L., Betzner A.S., Burn J., Witke W., Herth W.,			
RA	Camilleri C., Hoffe H., Plazinski J., Birch R., Cork A., Glover J.,			
RA	Redmond J., Williamson R.E.;			
RT	"Molecular analysis of cellulose biosynthesis in Arabidopsis.";			
RL	Science 279:17-20(1998).			
DR	EMBL: AF027174; AAC39336.1; "			
DR	InterPro: IPR005150; Cellulose_synt.			
DR	InterPro: IPR001841; znf_ring.			
DR	Pfam: PF03552; Cellulose_synt; 1.			
DR	SMART: SM00184; RING; 1.			
SO	SEQUENCE 1065 AA; 119659 MW; 1EF01C95FB85BE4C CRC64;			
Query Match				
Best Local Similarity 100.0%; Score 5677; DB 10; Length 1065;				
Matches 1065; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1 MSEGGETAGKPMKNIVPQTCICSDNNGKTVGDGFPAACDICSPPVCPCTCYERKDGNO 60			
DB	1 MSEGGETAGKPMKNIVPQTCICSDNNGKTVGDGFPAACDICSPPVCPCTCYERKDGNO 60			
OY	61 SCPCCKTRRYKRLKSPALPGKDEBDGLADECTVEFNPQKKEISERMGLMTGKGBEM 120			

```

Db      61 SCPOCKTRKRLKGSALIPGDDXDEDLAEGVIEFVYPOKEKISERMGLMHLRGGEEM 120
QY      121 GEPDYDEKSHHNLPLRTSRDTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
Db      121 GEPDYDEKSHHNLPLRTSRDTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
QY      181 IYDPVGLGNVAMKERVADGKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
Db      181 IYDPVGLGNVAMKERVADGKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
QY      241 ROPLSKRVSISSSRINPYRMVIMLRVITICLFLHYRITNPVPAFALMLVSYCEIEMFAL 300
Db      241 ROPLSKRVSISSSRINPYRMVIMLRVITICLFLHYRITNPVPAFALMLVSYCEIEMFAL 300
QY      301 SWILDOFPKMFVNNRTYIDRLALRDRGSEPSQLAAVDIFVSTVPLKEPPLVTANTYL 360
Db      301 SWILDOFPKMFVNNRTYIDRLALRDRGSEPSQLAAVDIFVSTVPLKEPPLVTANTYL 360
QY      361 SIILAVDPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEYFAAK 420
Db      361 SIILAVDPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEYFAAK 420
QY      421 IYLDKDVOTSEYKDRAMKREYEERKIRINMLVSKALCPREGWVMOGTPWPGNNTGD 480
Db      421 IYLDKDVOTSEYKDRAMKREYEERKIRINMLVSKALCPREGWVMOGTPWPGNNTGD 480
QY      481 HPGMIOVFLGONGGDAEGNELPRLYVYSREKRPQOHKKKAGAMALRVSAVLTNGPF 540
Db      481 HPGMIOVFLGONGGDAEGNELPRLYVYSREKRPQOHKKKAGAMALRVSAVLTNGPF 540
QY      541 IINLDCDHYINNSKALREAMCFMLDNLKQVCYVOPORFDGIDKNDRYANNTVEFDI 600
Db      541 IINLDCDHYINNSKALREAMCFMLDNLKQVCYVOPORFDGIDKNDRYANNTVEFDI 600
QY      601 NLRGIDGLOGPYVYGCVFNRTALYGPPIKVKRKKPSLSKLOGGSRKNSKAKKES 660
Db      601 NLRGIDGLOGPYVYGCVFNRTALYGPPIKVKRKKPSLSKLOGGSRKNSKAKKES 660
QY      661 DKKSGRHTDSTVPVFNLDIEGVSAGFDEKALMOMSLERFGOSAVVASTIME 720
Db      661 DKKSGRHTDSTVPVFNLDIEGVSAGFDEKALMOMSLERFGOSAVVASTIME 720
QY      721 NGGVPSATPENMLKEAHIVISGVEDKSDWGMETGIVSYTEDILTFKMHARGWRSI 780
Db      721 NGGVPSATPENMLKEAHIVISGVEDKSDWGMETGIVSYTEDILTFKMHARGWRSI 780
QY      781 YCMPRLPAFKGSAPINLSDRLNOVLWALGSVEILFSRHCPIMYGNGSLKFLERRAYVN 840
Db      781 YCMPRLPAFKGSAPINLSDRLNOVLWALGSVEILFSRHCPIMYGNGSLKFLERRAYVN 840
QY      841 TTYIPITSIPILMYCTLAACLTNOFIIPQISNINIASIMFLSLFSTIFATGILEMNSGV 900
Db      841 TTYIPITSIPILMYCTLAACLTNOFIIPQISNINIASIMFLSLFSTIFATGILEMNSGV 900
QY      901 GIDEMWRNQQFVYIGVSAHLFAVPOGILKVLADITNTVYSKASDEDEGAELYLFKN 960
Db      901 GIDEMWRNQQFVYIGVSAHLFAVPOGILKVLADITNTVYSKASDEDEGAELYLFKN 960
QY      961 TLLLPPTLLIVNLGVAVGSAVAINSGYOSGMPLEGLFAFWIVILVPLKGLMR 1020
Db      961 TLLLPPTLLIVNLGVAVGSAVAINSGYOSGMPLEGLFAFWIVILVPLKGLMR 1020
QY      1021 QNRTPTIVVMSVLASIFSLMVRIDPTSRVGTGDIIEGGINC 1065
Db      1021 QNRTPTIVVMSVLASIFSLMVRIDPTSRVGTGDIIEGGINC 1065

```

RESULT 2  
 Q9FHK6 PRELIMINARY; PRT: 1065 AA.  
 ID Q9FHK6  
 AC Q9FHK6  
 DT 01-MAR-2001 (Tremblrel. 16, Created)

```

DT      01-MAR-2001 (Tremblrel. 16, last sequence update)
DT      01-JUN-2002 (Tremblrel. 21, last annotation update)
DE      Cellulose synthase catalytic subunit.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxId=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=COLUMBIA;
RX      MEDLINE=99397451; PubMed=10470850;
RA      Kaneo T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA      Miyajima N., Tabata S.;
RT      "Structural analysis of Arabidopsis thaliana chromosome 5. IX
RT      Sequence features of the regions of 1,011,550 bp covered by seventeen
RT      P1 and TAC clones."
RL      DNA Res. 6:183-195(1999).
DR      EMBL: AB018111; BAB09693.1; -;
DR      InterPro: IPR005150; Cellulose_synth.
DR      InterPro: IPR001841; Znf_ring.
DR      Pfam: PF03552; Cellulose_synth; 1.
DR      SMART: SM00184; RING; 1.
SO      SEQUENCE 1065 AA; 119682 MW; 3AA4714CE3C4D581 CRC64;

Query Match      99.6%; Score 5654; DB 10; Length 1065;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1062; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 MSEGEGTAGKPMKNIVPOTCOICSDNVGKTVGDGRVYADICGFPYCRCEYERKDNQ 60
Db      1 MSEGEGTAGKPMKNIVPOTCOICSDNVGKTVGDGRVYADICGFPYCRCEYERKDNQ 60
QY      61 SCPOCKTRKRLKGSALIPGDDXDEDLAEGVIEFVYPOKEKISERMGLMHLRGGEEM 120
Db      61 SCPOCKTRKRLKGSALIPGDDXDEDLAEGVIEFVYPOKEKISERMGLMHLRGGEEM 120
QY      121 GEPDYDEKSHHNLPLRTSRDTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
Db      121 GEPDYDEKSHHNLPLRTSRDTSGEFSAASPERLSVSTIAGKRLPYSSDNOGPNRR 180
QY      181 IYDPVGLGNVAMKERVADGKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
Db      181 IYDPVGLGNVAMKERVADGKMKOEKNTGVSVOAASERGVDIDASTDILADEALLNDEA 240
QY      241 ROPLSKRVSISSSRINPYRMVIMLRVITICLFLHYRITNPVPAFALMLVSYCEIEMFAL 300
Db      241 ROPLSKRVSISSSRINPYRMVIMLRVITICLFLHYRITNPVPAFALMLVSYCEIEMFAL 300
QY      301 SWILDOFPKMFVNNRTYIDRLALRDRGSEPSQLAAVDIFVSTVPLKEPPLVTANTYL 360
Db      301 SWILDOFPKMFVNNRTYIDRLALRDRGSEPSQLAAVDIFVSTVPLKEPPLVTANTYL 360
QY      361 SIILAVDPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEYFAAK 420
Db      361 SIILAVDPVDKVSVCYFDDGAAMLSESLAETSEFARKWVPCKKYSIEPRABEYFAAK 420
QY      421 IYLDKDVOTSEYKDRAMKREYEERKIRINMLVSKALCPREGWVMOGTPWPGNNTGD 480
Db      421 IYLDKDVOTSEYKDRAMKREYEERKIRINMLVSKALCPREGWVMOGTPWPGNNTGD 480
QY      481 HPGMIOVFLGONGGDAEGNELPRLYVYSREKRPQOHKKKAGAMALRVSAVLTNGPF 540
Db      481 HPGMIOVFLGONGGDAEGNELPRLYVYSREKRPQOHKKKAGAMALRVSAVLTNGPF 540
QY      541 IINLDCDHYINNSKALREAMCFMLDNLKQVCYVOPORFDGIDKNDRYANNTVEFDI 600
Db      541 IINLDCDHYINNSKALREAMCFMLDNLKQVCYVOPORFDGIDKNDRYANNTVEFDI 600
QY      601 NLRGIDGLOGPYVYGCVFNRTALYGPPIKVKRKKPSLSKLOGGSRKNSKAKKES 660
Db      601 NLRGIDGLOGPYVYGCVFNRTALYGPPIKVKRKKPSLSKLOGGSRKNSKAKKES 660

```

QY 661 DKKSGRHTDSTVPVFNLDIEEGVAGAFDEKALMSOMSLERKFGOSAVFVASTIME 720  
 DB 661 DKKSGRHTDSTVPVFNLDIEEGVAGAFDEKALMSOMSLERKFGOSAVFVASTIME 720  
 QY 721 NGCVPPSPATPNLKEAIVHISCGYEDKSDMGMEIGWYGVTEEDILTGFKMHARGMRSI 780  
 DB 721 NGCVPPSPATPNLKEAIVHISCGYEDKSDMGMEIGWYGVTEEDILTGFKMHARGMRSI 780  
 QY 781 YCMKPLPAFGSAPINLSDRINOVYRMALGSVEILFHSRCPHWGYNRKLFRLEFAVYN 840  
 DB 781 YCMKPLPAFGSAPINLSDRINOVYRMALGSVEILFHSRCPHWGYNRKLFRLEFAVYN 840  
 QY 841 TTIVPTISIPILMYCTLLAVCLFTNOFTIIPQISNIAISIFLSLFIATGILERMMSGV 900  
 DB 841 TTIVPTISIPILMYCTLLAVCLFTNOFTIIPQISNIAISIFLSLFIATGILERMMSGV 900  
 QY 901 GIDEMWNRQFVWYIGVSAHLFAVFOGILKVLAGIDTNTFYTSKASDEGDFAEILYFKW 960  
 DB 901 GIDEMWNRQFVWYIGVSAHLFAVFOGILKVLAGIDTNTFYTSKASDEGDFAEILYFKW 960  
 QY 961 TTLILPPTLLILVNVGVVAGSVYAINSGYSGMPLFKPLFAFVIVHLVPELGLMGR 1020  
 DB 961 TTLILPPTLLILVNVGVVAGSVYAINSGYSGMPLFKPLFAFVIVHLVPELGLMGR 1020  
 QY 1021 QNRPTIYVWVSVLLASIFSLIMVRIIDPFTSRVTGPDIIEGGINC 1065  
 DB 1021 QNRPTIYVWVSVLLASIFSLIMVRIIDPFTSRVTGPDIIEGGINC 1065

## RESULT 3

QY 09XG6 PRELIMINARY: PRT: 1067 AA.  
 AC 09XG6:  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Cellulose synthase catalytic subunit.  
 GN CEL3.  
 OS Gossypium hirsutum (Upland cotton).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Malvales; Malvaceae; Gossypium.  
 OX NCBI\_TaxID=3635;  
 RX [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=CV. TEXAS MARKER-1; TISSUE=COTTON FIBER;  
 RX MEDLINE=20098737; PubMed=10631273;  
 RA laosincha1 W.; Cui X.; Brown R.M. JR.;  
 RT "A full length cDNA of cotton cellulose synthase has high homology  
 with the Arabidopsis RSM1 gene and the cotton Cel1 gene (Accession  
 No. AF200453) (PCR 00-002).";  
 RL Plant Physiol. 122:291-291(2000).  
 RN [2]  
 RN SEQUENCE FROM N. A.  
 RC STRAIN=CV. TEXAS MARKER-1; TISSUE=COTTON FIBER;  
 RA Kimura S.; laosincha1 W.; Itoh T.; Cui X.; Brown R.M. JR.;  
 RT "Immunogold labeling of Rosette Terminal Cellulose Synthetizing  
 RT Complexes in a Vascular Plant (Vigna angulatis).";  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF150630; AAD9534.2; -  
 DR InterPro: IPR005150; Cellulose\_synth.  
 DR InterPro: IPR001841; znf\_ring.  
 DR Pfam: PF03552; Cellulose\_synth; 1.  
 DR SMART; SM00184; RING; 1.  
 SQ SEQUENCE 1067 AA; 119325 MW; 0BBA2ED00590F29C CRC64;

Query Match 87.3%; Score 4954; DB 10; Length 1067;  
 Best Local Similarity 85.3%; Pred. No. 0;  
 Matches 919; Conservative 75; Mismatches 60; Indels 24; Gaps 6;

QY 1 MESGEAGAKPMKNIIVPTQICSDNVGKTVGDRFVACDICSFFVRCRCYERKDGNO 60  
 DB 1 MESEGDIGCKPMKNIGGQTCICGDNVGNKTDGDFIACNCAFCVRCRCYERKDGNO 60

QY 61 SCPOCKTRKRLKSGPALPGDKEDGLADEGTFVE-----NYPQEKISERMLGHLTRGK 116  
 DB 61 SCPOCKTRKRLKSGPALPGDKEDGLADEGTFVE-----NYPQEKISERMLGHLTRGK 116  
 QY 117 GEEMGEPOYDKESVSHNHLPRLTSRQDTSGEPSAASPERLSVST-IAGGKRLPYSSDVNO 175  
 DB 117 GEEMGEPOYDKESVSHNHLPRLTSRQDTSGEPSAASPERLSVST-IAGGKRLPYSSDVNO 175  
 QY 121 GEDVGAFTYDKETSHNHIFPLTISGQEVSGELSAASPERLSMASPGVAGK----- 170  
 DB 121 GEDVGAFTYDKETSHNHIFPLTISGQEVSGELSAASPERLSMASPGVAGK----- 170  
 QY 176 SPNRKRIIDVP-----GIGNNAKFRVNGKMKKOKNNGPVST-QAASRGVDIDASTD 228  
 DB 176 SPNRKRIIDVP-----GIGNNAKFRVNGKMKKOKNNGPVST-QAASRGVDIDASTD 228  
 QY 229 ILADEALLNDEAROPLSRKSVPSSIRIPYRWILRIYILCLFPHYRTNVPNAPALM 288  
 DB 229 ILADEALLNDEAROPLSRKSVPSSIRIPYRWILRIYILCLFPHYRTNVPNAPALM 288  
 QY 230 VLVDSQNLNDEAROPLSRKSVPSSIRIPYRWILRIYILCLFPHYRTNVPNAPALM 289  
 DB 230 VLVDSQNLNDEAROPLSRKSVPSSIRIPYRWILRIYILCLFPHYRTNVPNAPALM 289  
 QY 289 LVSVACEIWFALSWLIDQFPKWFVNNRETYLDRLALRYDREGEPSQLAAVDIFSVTDPL 348  
 DB 289 LVSVACEIWFALSWLIDQFPKWFVNNRETYLDRLALRYDREGEPSQLAAVDIFSVTDPL 348  
 QY 290 LISVACEIWFALSWLIDQFPKWFVNNRETYLDRLALRYDREGEPSQLAAVDIFSVTDPL 349  
 DB 290 LISVACEIWFALSWLIDQFPKWFVNNRETYLDRLALRYDREGEPSQLAAVDIFSVTDPL 349  
 QY 349 KEPPLVTANTVLSILAVDPVDKVCYFDDGAMLSFESLAEISFAPKRVYFCKKYSI 408  
 DB 349 KEPPLVTANTVLSILAVDPVDKVCYFDDGAMLSFESLAEISFAPKRVYFCKKYSI 408  
 QY 350 KEPPLVTANTVLSILAVDPVDKVCYFDDGAMLSFESLAEISFAPKRVYFCKKYSI 409  
 DB 350 KEPPLVTANTVLSILAVDPVDKVCYFDDGAMLSFESLAEISFAPKRVYFCKKYSI 409  
 QY 409 EPRAPEMYFAAKIDYLDKQVOTSFVKDRAMKREYEFKIRINALVSKALCPREGVWMO 468  
 DB 409 EPRAPEMYFAAKIDYLDKQVOTSFVKDRAMKREYEFKIRINALVSKALCPREGVWMO 468  
 QY 410 EPRAPEMYFAAKIDYLDKQVOTSFVKDRAMKREYEFKIRINALVSKALCPREGVWMO 469  
 DB 410 EPRAPEMYFAAKIDYLDKQVOTSFVKDRAMKREYEFKIRINALVSKALCPREGVWMO 469  
 QY 469 DGTWPGNNTGDHPGMIQVFLGONGDLAEGNELRLVYVSEKRRPGQHHKAGAMNL 528  
 DB 469 DGTWPGNNTGDHPGMIQVFLGONGDLAEGNELRLVYVSEKRRPGQHHKAGAMNL 528  
 QY 470 DGTWPGNNTGDHPGMIQVFLGONGDLAEGNELRLVYVSEKRRPGQHHKAGAMNL 529  
 DB 470 DGTWPGNNTGDHPGMIQVFLGONGDLAEGNELRLVYVSEKRRPGQHHKAGAMNL 529  
 QY 529 VVSVAVLTNGPFTLMDCDHYTNSKALREACPLMDPDLKQVCYVQFQRFQDIDKND 588  
 DB 529 VVSVAVLTNGPFTLMDCDHYTNSKALREACPLMDPDLKQVCYVQFQRFQDIDKND 588  
 QY 530 VVSVAVLTNGPFTLMDCDHYTNSKALREACPLMDPDLKQVCYVQFQRFQDIDKND 589  
 DB 530 VVSVAVLTNGPFTLMDCDHYTNSKALREACPLMDPDLKQVCYVQFQRFQDIDKND 589  
 QY 589 RYANNTVPEFDINLGLDGIQGPVYVGVGVNPRALCYEPRIVKHKHKKPILLSKLCG 648  
 DB 589 RYANNTVPEFDINLGLDGIQGPVYVGVGVNPRALCYEPRIVKHKHKKPILLSKLCG 648  
 QY 590 RYANNTVPEFDINLGLDGIQGPVYVGVGVNPRALCYEPRIVKHKHKKPILLSKLCG 649  
 DB 590 RYANNTVPEFDINLGLDGIQGPVYVGVGVNPRALCYEPRIVKHKHKKPILLSKLCG 649  
 QY 649 SRKKNKSK-AKRESDDKKSGRHTDSTVPVFNLDIEEGVAGAFDEKALMSOMSLERK 707  
 DB 649 SRKKNKSK-AKRESDDKKSGRHTDSTVPVFNLDIEEGVAGAFDEKALMSOMSLERK 707  
 QY 650 SRKKNKSKSKKSDKKKSGKHVDSTVPVFNLDIEEGVAGAFDEKALMSOMSLERK 709  
 DB 650 SRKKNKSKSKKSDKKKSGKHVDSTVPVFNLDIEEGVAGAFDEKALMSOMSLERK 709  
 QY 708 GOSAVFVASTIMENGVPPSPATPNLKEAIVHISCGYEDKSDMGMEIGWYGVTEEDIL 767  
 DB 708 GOSAVFVASTIMENGVPPSPATPNLKEAIVHISCGYEDKSDMGMEIGWYGVTEEDIL 767  
 QY 710 GOSAVFVASTIMENGVPPSPATPNLKEAIVHISCGYEDKSDMGMEIGWYGVTEEDIL 769  
 DB 710 GOSAVFVASTIMENGVPPSPATPNLKEAIVHISCGYEDKSDMGMEIGWYGVTEEDIL 769  
 QY 768 TGFKNHARGMRSIYCMKPLPAFGSAPINLSDRINOVYRMALGSVEILFHSRCPHWGYN 827  
 DB 768 TGFKNHARGMRSIYCMKPLPAFGSAPINLSDRINOVYRMALGSVEILFHSRCPHWGYN 827  
 QY 828 GRLKFLERFAVYNTIYPTISIPILMYCTLLAVCLFTNOFTIIPQISNIAISIFLSLFI 887  
 DB 828 GRLKFLERFAVYNTIYPTISIPILMYCTLLAVCLFTNOFTIIPQISNIAISIFLSLFI 887  
 QY 830 GRLKFLERFAVYNTIYPTISIPILMYCTLLAVCLFTNOFTIIPQISNIAISIFLSLFI 889  
 DB 830 GRLKFLERFAVYNTIYPTISIPILMYCTLLAVCLFTNOFTIIPQISNIAISIFLSLFI 889  
 QY 888 FATGILERMMSGVIGIDEMWNRQFVWYIGVSAHLFAVFOGILKVLAGIDTNTFYTSKASD 947  
 DB 888 FATGILERMMSGVIGIDEMWNRQFVWYIGVSAHLFAVFOGILKVLAGIDTNTFYTSKASD 947  
 QY 890 FATGILERMMSGVIGIDEMWNRQFVWYIGVSAHLFAVFOGILKVLAGIDTNTFYTSKASD 949  
 DB 890 FATGILERMMSGVIGIDEMWNRQFVWYIGVSAHLFAVFOGILKVLAGIDTNTFYTSKASD 949  
 QY 948 EDGDFAEILYFKWTTLLIPPTLLILVNVGVVAGSVYAINSGYSGMPLFKPLFAFVIV 1007  
 DB 948 EDGDFAEILYFKWTTLLIPPTLLILVNVGVVAGSVYAINSGYSGMPLFKPLFAFVIV 1007  
 QY 950 EDGDFAEILYFKWTTLLIPPTLLILVNVGVVAGSVYAINSGYSGMPLFKPLFAFVIV 1009  
 DB 950 EDGDFAEILYFKWTTLLIPPTLLILVNVGVVAGSVYAINSGYSGMPLFKPLFAFVIV 1009

RESULT 4  
 QY 99LLI6 PRELIMINARY: PRT: 1077 AA.  
 ID 99LLI6



```

Db      62  PCKNNKYNKHKSGPAIRGEEDDTDADAS--DFNYPASGNDQOKIADRMKSWRNAG 120
Qy      117  GEEMGEPOYDK-----EVSHNHLRLTSRODNGSEPSAASPERLSYSTIAGK 165
Db      121  SCDVGRPKTDSGEITLTYDSEGLRGTIPSYTNSO--ISGELPGASPDHMSPTGNICR 179
Qy      166  RLPS--SDVNSPNRRIYDPVGLGNVAMKERVDGKMKOEKNTGPV---STOASERGV 221
Db      180  RAPEFYMHMSSNPSREFSGV--GNVAMKERVDGKMKODKKTIPNTNTSTAPSGRGV 237
Qy      222  -DIDASTDLADEALLNDEAROPLSRKVSIPSSRINPYRMVIMRLVILCLFHYRTNP 280
Db      238  GDIDASTDYNMEDALLNDETRQPLSRKVPPLPSSRINPYRMVIVLRLIVLSTFLHYRTNP 297
Qy      281  VPNAFLMLVSYICIMFALSWILDOFPKMPVPNRETYSRLALRDREGESQLAADV 340
Db      298  VRNAFLPLMLVSYICIMFALSWILDOFPKMPVPNRETYSRLALRDREGESQLAADV 357
Qy      341  FVSTVDPLKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLSFESLAESEFARKV 400
Db      358  FVSTVDPMKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLSFESLAESEFARKV 417
Qy      401  PECKKSYIEPRAPERYFAKIDYLYKQVQTSVKNDRANKREYEERKIRINALVSKALC 460
Db      438  PEVKRYNIEPRAPERYFSQIDYLYKDKVHPSEVKDRRAKREYEERKIRYNGLVAKQV 477
Qy      461  PEEGVNMDGTPWPNNGDHPGMIQVLFQNGSGIDAEENELRLVYVSREKRPQOHK 520
Db      478  PEEGVNMDGTPWPNNTRDHPGMIQVLFQNGSGIDAEENELRLVYVSREKRPQOHK 537
Qy      521  KAGANALVRSVAVLTNGPFILNLCDDHYINNSKALREACFLMPNLKQVQYQFOPR 580
Db      538  KAGANALVRSVAVLTNGQYMLNLCDDHYINNSKALREACFLMPNLKQVQYQFOPR 597
Qy      581  FDGIDKNDRIANRNTVFEDINLRGLDIOGPVYVGTGCVFNRTALYGPPIKVKHKPS 640
Db      598  FDGIDRNDRIANRNTVFEDINLRGLDIOGPVYVGTGCVFNRTALYGPPI--KQKKG 655
Qy      641  LLSKLCGSRKKNSKAKKESDCKGGRHDSIVPVPNLDDIEGVGAGFDEKSLMSQ 700
Db      656  FLSSLCGG--RKGSGSKSKGSDKSKQKQKHYDSSVPVFNEDIEGVGAGFDEKSLMSQ 714
Qy      701  MSLERFGQSAFVASTLMENGVPPSATPENILKEAIIHIVISGVEDKSDMGMEIGW 760
Db      715  MSLERFGQSAFVASTLMETGVQPSATPESLKEAIIHIVISGVEDKTEMGETIWMG 774
Qy      761  SYTEDILITGFKMHAGMSIYCMKPLPAKGSAPINLSDRLNOVLKMGALGSVEILFSRH 820
Db      775  SYTEDILITGFKMHAGMSIYCMKPRPAKGSAPINLSDRLNOVLKMGALGSVEILFSRH 834
Qy      821  PIWYGYNGRLKFLERPAVNTIYPTISPLMYCGLAVCLETNOFIROIISNASTMF 880
Db      835  PIWYGYNGRLKFLERPAVNTIYPTISPLMYCGLAVCLETNOFIROIISNASTMF 894
Qy      881  LSLFLSIFATGILEKMSVGVIGDEMWRNEQFVAVIGVSAHLEFAVEGILKVLADITNP 940
Db      895  LSLFLSIFATGILEKMSVGVIGDEMWRNEQFVAVIGVSAHLEFAVEGILKVLADITNP 954
Qy      941  VTSKASDEBDGDAELYLKWTLLIPPTLLIVNLVGVAVSYAINSGYSGMPLFGL 1000
Db      955  VTSKASDEBDGDAELYLKWTLLIPPTLLIVNLVGVAVSYAINSGYSGMPLFGL 1014
Qy      1001  FFAFVYIHLVPELGLMGROKRPPIYVVMVSLASTISLWVRIDPTSVTGDILE 1060
Db      1015  FFAFVYIHLVPELGLMGROKRPPIYVVMVSLASTISLWVRIDPTSVTGDILE 1074
Qy      1061  CGINC 1065
Db      1075  CGINC 1079

```

RESULT 6  
Q9LL15

```

ID      Q9LL15  PRELIMINARY;  PRT;  1076 AA.
AC      Q9LL15;
DT      01-OCT-2000 (Tremblrel. 15, Created)
DT      01-OCT-2000 (Tremblrel. 15, last sequence update)
DT      01-JUN-2002 (Tremblrel. 21, last annotation update)
DE      Cellulose synthase-5.
GN      CESA-5.
OS      Zea mays (maize).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACC clade;
OC      Panicoideae; Andropogoneae; Zea.
OX      NCBI_TaxID=4577;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20398328; PubMed=10938350;
RA      Holland N., Holland D., Helentjaris T., Dhugga K.S.,
RA      Xocoostle-Cazares B., Delmer D.P.;
RT      "A comparative analysis of the plant cellulose synthase (Cesa) gene
RT      family";
RL      Plant Physiol. 123:1313-1324(2000).
DR      EMBL: AF200529; AAF89965.1;
DR      InterPro: IPR005150; Cellulose_synth.
DR      InterPro: IPR001841; Znf_ring.
DR      Pfam: PF03552; Cellulose_synth; 1.
DR      SMART: SM00184; RING; 1.
SQ      SEQUENCE 1076 AA; 120724 MW; 41E7C20EDA54F27 CRC64;

```

Query Match 79.8%; Score 4529.5; DB 10; Length 1076;

Best Local Similarity 78.1%; Pred. No. 0;

Matches 847; Conservative 88; Mismatches 118; Indels 31; Gaps 10;

```

Qy      4  EEFACKPKKNIVPQICSDNVGKTVGDRFVACDIFPCRCPCYERKDNQSP 63
Db      2  DGDATNSCKHVAQVQCICGDSVGTADGDLFTACDVGCFPCRCCTEYERKDGQACP 61
Qy      64  OCTKRYRLKGSALPFGDDEGLADEGVEFNY-----POKEKISEBMLNHLTRGK 117
Db      62  OCTKRYKHKSGSPVHGEENEVDADD--VSDNYQASGNQDQOKIAEMLTMR--TNSG 119
Qy      118  EEMGEPOYDK-----EVSHNHLRLTSRODTSCGFSAASPERL--SVSSTIAG 164
Db      120  SDIGLAKYDSGEIGHGKYDSGEIPRGYISLTHSQ--ISGELPGASPDHMSPTGNICR 178
Qy      165  KRLPSDDVNSPNRRIYDPVGLGNVAMKERVDGKMKOE--KMGVPTSOASRGV 221
Db      179  HOPPY--VNSHNSRERFSGSLGNVAMKERVDGKMKOGAIPTMGTSLAPSGRGV 235
Qy      222  DIDASTDLADEALLNDEAROPLSRKVSIPSSRINPYRMVIMRLVILCLFHYRTNP 281
Db      236  DIDASTDYNMEDALLNDETRQPLSRKVPPLPSSRINPYRMVIVLRLIVLSTFLHYRTNP 295
Qy      282  PNAFLMLVSYICIMFALSWILDOFPKMPVPNRETYSRLALRDREGESQLAADV 341
Db      296  PNAFLMLVSYICIMFALSWILDOFPKMPVPNRETYSRLALRDREGESQLAADV 355
Qy      342  VSTVDPLKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLSFESLAESEFARKV 401
Db      356  VSTVDPMKEPPLVANTVLSILAVDYPVKVSCYVDDGAMLSFESLAESEFARKV 415
Qy      402  PECKKSYIEPRAPERYFAKIDYLYKQVQTSVKNDRANKREYEERKIRINALVSKALC 461
Db      416  PECKKSYIEPRAPERYFAKIDYLYKQVQTSVKNDRANKREYEERKIRYNGLVAKQV 475
Qy      462  EBGVNMDDGTPWPNNGDHPGMIQVLFQNGSGIDAEENELRLVYVSREKRPQOHK 521
Db      476  EBGVNMDDGTPWPNNTRDHPGMIQVLFQNGSGIDAEENELRLVYVSREKRPQOHK 535
Qy      522  AGAMNALVRSVAVLTNGPFILNLCDDHYINNSKALREACFLMPNLKQVQYQFOPR 581
Db      536  AGAMNALVRSVAVLTNGQYMLNLCDDHYINNSKALREACFLMPNLKQVQYQFOPR 595
Qy      582  DGIDKNDRIANRNTVFEDINLRGLDIOGPVYVGTGCVFNRTALYGPPIKVKHKPS 641

```

```

Db      596 DGIIDRNDRIANRNTVFEDINLRLGLDGIQGFVYVTCGVENRTALGYEPP--VKKKKPGF 653
Qy      642 LSKICGSGRRKKNKAKKESDCKKSGRHTSTVPFNLDIEEGVEGAGFDEKALLMSQM 701
Db      654 FSSLGCG--RKKTSKSKSSKSKSHRHAOSVFPNLEDIEEGIEGQFDEKSLMSQM 712
Qy      702 SLEKRRGQSAVFPAVSLMENGVPSPATPENLKEAIVHISGCEYEDKSDMGMEIGWYGS 761
Db      713 SLEKRGQSSVFPAVSLMEYGVPOSATPELSLKEAIVHISGCEYEDKSDMGMEIGWYGS 772
Qy      762 VTEIDITLGFPMHARGRSIYCMPLKAPGSAFINLSDRLNOYLRLMALGSVELLESHCP 821
Db      773 VTEIDITLGFPMHARGRSIYCMPLKAPGSAFINLSDRLNOYLRLMALGSVELLESHCP 832
Qy      822 IMVNGRRLKLEFRFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISINIASIWL 881
Db      833 IMVNGRRLKLEFRFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISINIASIWL 892
Qy      882 SLEFLSIPATGILEMRNSGVIDEMWRNEOFWVIGVSAHLFAVQGLKVLAVGIDITNFTV 941
Db      893 SLEFLSIPATGILEMRNSGVIDEMWRNEOFWVIGVSAHLFAVQGLKVLAVGIDITNFTV 952
Qy      942 TSASAPEDGDEFAELTYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSYGQSMGPLFGKLF 1001
Db      953 TSASAPEDGDEFAELTYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSYGQSMGPLFGKLF 1012
Qy      1002 FAFWVIVHLYPFLKLGMRONRPTIYVWVSVLASIFSLMVRIDPETSRYGPDILEC 1061
Db      1013 FAFWVIVHLYPFLKLGMRONRPTIYVWVSVLASIFSLMVRIDPETSRYGPDILEC 1072
Qy      1062 GINC 1065
Db      1073 GINC 1076

```

RESULT 7  
09LL19 PRELIMINARY; PRT; 1075 AA.

Query Match 71.0%; Score 4028.5; DB 10; Length 1075;  
Best Local Similarity 69.6%; Pred. No. 4e-304;  
Matches 752; Conservative 130; Mismatches 153; Indels 45; Gaps 12;

```

Qy      4 EGETAG--KPMKNIVPOTCOICSDNKGTVGDRFVACICISHPVPCPEYERKDGNO 61
Db      23 DGDVPSAKPTKSANGVQICGSDGVSATGTVFACMECAPVPCPEYERKDGNO 82
Qy      62 CPCKTRRYRLKAGSPALPDCKDEGLADGVEYENYPOKERSRMGLMHLTRGKEEMG 121

```

```

Db      83 CPCKTRRYRKOKSPRVHGDDEDEDV--DDIDNEFNKOGS-----GKGPWMO 128
Qy      122 EPQYDKVS-----HNHLRLTSDOTSGEASAPESLVSSTIAGKRIPYSVDN 174
Db      129 LOGDDADLSSAHEPHHRIRPLRTSGOQISGELIPDASPDRHSI-----RPTSSYD 180
Qy      175 QS--PNRRIVDP-----VGLGNVANKERYDGMKMOEKNTPGVSQAASERGVVIDAS 226
Db      181 PSVPVPRIVDPKDLNLSYGLNSGVNDEKERVESMRVRODKMMQVTKNYPARGG--DME-G 238
Qy      227 TDLADEALLNDAROPLSRKVSIPSSRIPIYVIMLRLVITCLELHRITNPNRANA 266
Db      239 TGSNGEMQWAVDARPLSLRIVPISNQLNLYRVVILRLILICFFQYRVSHVVDAYG 298
Qy      287 LMLVSVICELWFLSMILDOFPKMPFNRETYIDRLALRDREGEQSOLAVIDFSTVD 346
Db      299 LMLVSVICELWFLSMILDOFPKMPFNRETYIDRLALRDREGEQSOLAVIDFSTVD 358
Qy      347 PLKEPPLVANTVLSILAVDYPVDKVCYVFDGGAAMLSESLAETSEFARKVPPCKRY 406
Db      359 PLKEPPLVANTVLSILAVDYPVDKVCYVFDGGAAMLSESLAETSEFARKVPPCKRY 418
Qy      407 SIEPRAPWYFAAKIDYLDKQVOTSFYKDRRAKREYEEKIRINALVSKALCPDEGWY 466
Db      419 NIEPRAPWYFAAKIDYLDKQVOTSFYKDRRAKREYEEKIRINALVSKALCPDEGWY 478
Qy      467 MDSGTWPGNNTDHDHGMIOVFLGONGIDAENELPRLYVSRERKPGOHKKKGAAM 526
Db      479 MDSGTWPGNNTDHDHGMIOVFLGONGIDAENELPRLYVSRERKPGOHKKKGAAM 538
Qy      527 ALTVASAVLNGPFIILNCDHYINNSKALREAMCELMDPNLKQVCYVOFPRPFGIDK 586
Db      539 ALTVASAVLNGPFIILNCDHYINNSKALREAMCELMDPNLKQVCYVOFPRPFGIDK 598
Qy      587 NDRYANRNTVFPDINLRGLDIOGPYVGTGCVFNRTALGYEPPIKVKHKKSLSKLC 646
Db      599 HDRYANRNTVFPDINLRGLDIOGPYVGTGCVFNRTALGYEPPIKVKHKKSLSKLC 658
Qy      647 GGSRRKKNKSKAKKESDCKKSGRHTSTVPFNLDIEEGVEGAGFDEKALLMSQM 706
Db      659 GGSRRKKNKSKAKKESDCKKSGRHTSTVPFNLDIEEGVEGAGFDEKALLMSQM 715
Qy      707 FGQSAVFVASTLMENGVPSPATPENLKEAIVHISGCEYEDKSDMGMEIGWYGSYTEDI 766
Db      716 FGQSAVFVASTLMENGVPSPATPENLKEAIVHISGCEYEDKSDMGMEIGWYGSYTEDI 775
Qy      767 LTFGFKMHARGMSIYCMPLKAPGSAFINLSDRLNOYLRLMALGSVELLESHCP 826
Db      776 LTFGFKMHARGMSIYCMPLKAPGSAFINLSDRLNOYLRLMALGSVELLESHCP 835
Qy      827 NGRLKLEFRFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISINIASIWL 886
Db      836 NGRLKLEFRFAYVNTIYPTISIPLLMYCTLLAVCLFTNOFIPOISINIASIWL 895
Qy      887 IFATGILEMRNSGVIDEMWRNEOFWVIGVSAHLFAVQGLKVLAVGIDITNFTV 946
Db      896 IFATGILEMRNSGVIDEMWRNEOFWVIGVSAHLFAVQGLKVLAVGIDITNFTV 955
Qy      947 DEDGDEFAELTYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSYGQSMGPLFGKLF 1006
Db      956 DEDGDEFAELTYEFKMTTLLIPPTLLIIVNLGVVAGVSAINSYGQSMGPLFGKLF 1015
Qy      1007 IVHLYPFLKLGMRONRPTIYVWVSVLASIFSLMVRIDPETSRYGPDILEC 1065
Db      1016 IVHLYPFLKLGMRONRPTIYVWVSVLASIFSLMVRIDPETSRYGPDILEC 1075

```

RESULT 8  
09LL18 PRELIMINARY; PRT; 1074 AA.  
09LL18, 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)  
OC Cellulose synthase-2.  
OS CESA-2.  
GN Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoideae; Andropogoneae; Zea.  
ON NCBI\_TaxID=4577;  
RP SEQUENCE FROM N.A.  
RA MEDLINE=20398328; PubMed=10938350;  
RA Holland N., Holland D., Helentjaris T., Dhugga K.S.,  
RA Xocoostle-Cazares B., Delmer D.P.;  
RT "A comparative analysis of the plant cellulose synthase (cesa) gene  
RT family";  
RT Plant Physiol. 123:1313-1324(2000).  
DR EMBL: AF200526; AAF89962.1; -;  
DR InterPro: IPR001510; Cellulose\_synth.  
DR InterPro: IPR001841; znf\_ring.  
DR Pfam: PF03552; Cellulose\_synth. 1.  
DR SMART: SM00184; RING: 1.  
SQ SEQUENCE 1074 AA; 121187 MW; A11666F0564E210 CRC64;

Query Match 70.9%; Score 4024; DB 10; Length 1074;  
Best Local Similarity 69.4%; Pred. No. 9e-304;  
Matches 749; Conservative 128; Mismatches 159; Indels 44; Gaps 10;

QY 2 ESEGEFAGKPKMKNIYPTQICSDNKGTVGDRFVACDICSFPCPCYEYERKDNOS 61  
DB 23 DGGAPVPAKPTKSAAGVOCICGDTVGVSATGDFVACNCAFPVCRPCYEYERKDNOC 82  
QY 62 CPQCKTRKRLKSPALPGDKDEGLADECTVEFNPKQKISERNMGLTNGKEEMG 121  
DB 83 CPQCKTRKRLKSPALPGDKDEGLADECTVEFNPKQKISERNMGLTNGKEEMG 128  
QY 122 EPQYDKEVS-----HNHLPLRLTSRODTSGEFSASPERLSYSTAGGRPLYSVDN 174  
DB 129 LQGDADLSSARHDPHRRPLRLTSGQISGEIRPASPDSHST-----RSPTSYVD 180  
QY 175 QS--PKRRIYD-----VGLGVANKERYDGNKMKOEKKTGVSQTQASERGVDDIAS 226  
DB 181 PSYVPEVRIYDPSKDLNSYGLNSVDWKERYESWRVXODKMKMLQVTNKYPRARCDME--G 237  
QY 227 TDLADLALNDAROLPSKRSVIPSRIINRYMVIIMRLVITCLFHYETRLPYPAFA 286  
DB 238 TSGNGEDMOWVDARLPLSRIVYISSQMLNRYIILRLILLCFFQYRISHPRVAY 297  
QY 287 LMLVSYICEIWFALSWILDQFPKMFVNNRETYDLRLALRYDRGERSQLAAVDIFVSTD 346  
DB 298 LMLVSYICEVWFALSWILDQFPKMFVNNRETYDLRLALRYDRGERSQLAVIDVSTD 357  
QY 347 PLKEPPLVYANVYLSLANDYPVDKVSCYVFDGAMLSFESLAETSEFARKKVPCKKY 406  
DB 358 PLKEPPLVYANVYLSLANDYPVDKVSCYVFDGAMLSFESLAETSEFARKKVPCKKH 417  
QY 407 SIERRAPENYFAKIDYLYDKQVOTSFYKDRBRAMKREYEKKITNMLYSALCPRECVW 466  
DB 418 NIERRAPENYFAKIDYLYDKQVOTSFYKDRBRAMKREYEKKITNMLVAAQVPEEGWT 477  
QY 467 MODGTPEPNNNTDHPGMIQVFLGONGGLDAEGNELPRLVYVSREKRPFGOHKKAGAM 526  
DB 478 MADGTAMPGNPNRDPHGMIOVFLGHSGGLDTGDNELPRLVYVSREKRPFGOHKKAGAM 537  
QY 527 ALVRSVAVLTNGPFIINLDCDHYINNSKALREAMCLMDNLCQVYVFPORFGIDK 586  
DB 538 ALVRSVAVLTNGAYLLNVDCDHYNSKALREAMCMMDALRKTCYVFPORFGIDL 597  
QY 587 NDRYARNYFEPIINLGLDGTGGPYVGNHGFNFATLYGYPPIKVKKKKSLSLKIC 646  
DB 598 HDRYARNYFEPIINLGLDGTGGPYVGNHGFNFATLYGYPPIKVKKKKSLSLKIC 657  
QY 647 GGRKRNKSKAKESDKKSGRHTDSTVVFNLDDIEGVAGAFDEKALLMSQMSLEKR 706

DB 658 CGRRKRNKSYMDSQSRIMK-TESSAPIFNMEDIEGIE--GYEDERSVLMGQRKLEKR 714  
QY 707 FGOSAVFVASTLWENGVPSPASATPENILKRAIHVISCYGDKSDMGEIMYIGSVTEDI 766  
DB 715 FGOSPIFIATFTMGIPSTNPASTLKEAHIVHISCYGDKEWGEIMYIGSVTEDI 774  
QY 767 LTGFKMHARGMSRISYCPKLPFAKGSAPINLSRNLNOVLMWAGSYIELFSRRCPYIGY 826  
DB 775 LTGFKMHARGMSRISYCPKLPFAKGSAPINLSRNLNOVLMWAGSYIELFSRRCPYIGY 834  
QY 827 NGRKLEPFAVNTYPTISIPLMYCTPLAVCETNPTIPQISINIASIFLS 886  
DB 835 NGRKLEPFAVNTYPTISIPLMYCTPLAVCETNPTIPQISINIASIFLS 894  
QY 887 IFATGILEMWSGVGIDEMRNQFVYIGVSAHLPAFVGIGILKVLADITNTVYSKAS 946  
DB 895 IFATGILEMWSGVGIDEMRNQFVYIGVSAHLPAFVGIGILKVLADITNTVYSKAS 954  
QY 947 DEGDPAELFKMTLLIPTLLVNLGVAGVAYAINSGVSGPLFGKLPFAFV 1006  
DB 955 DEGDPAELFKMTLLIPTLLVNLGVAGVAYAINSGVSGPLFGKLPFAFV 1014  
QY 1007 IVHLYPEFLKGLMGKRONPTPIVYVSVLASIFSLSLWVRIDPFTSRVTPDIL-ECGINC 1065  
DB 1015 ILHLYPEFLKGLMGKRONPTPIVYVSVLASIFSLSLWVRIDPFTSRVTPDIL-ECGINC 1074

RESULT 9  
O48946 PRELIMINARY; PRT; 1081 AA.  
ID O48946;  
AC O48946;  
DE 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, last sequence update)  
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)  
DE Cellulose synthase catalytic subunit.  
GN RSM1 OR F884.110 OR AT4G32410.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
OC NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=98111412; PubMed=9445479;  
RA Atiolli T., Peng L., Beltzner A.S., Burn J., Wiltke W., Herth W.,  
RA Camilleri C., Hofte H., Plazinski J., Birch R., Cork A., Glover J.,  
RA Redmond J., Williamson R.E.;  
RT "Molecular analysis of cellulose biosynthesis in Arabidopsis.";  
RT Science 279:717-720(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Terry N., Ardiles W., Buyschaert C., Dasseville R.,  
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,  
RA Villarroel R., Gleien J., Van Montagu M., Hohelsel J., Mewes H.W.,  
RA Meyer K.F.X., Lemcke K., Schueller C.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Terry N., Ardiles W., Buyschaert C., Dasseville R., De Clerck R.,  
RA De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarroel R.,  
RA Gleien J., Van Montagu M., Mewes H.W., Lemcke K., Meyer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF027172; AAC39334.1; -;

Query Match	70.2%	Score 3984	DB 10	Length 1081
Best Local Similarity	69.7%	Pred. No. 1.2e-300		
Matches 741	Conservative 128	Mismatches 158	Indels 36	Gaps 13
OY	2	ESEGETACKPMKNVIPOFCOICSDNVDGTYVDGDRFVADDCISFPVCRPCRYEERKDDGNS	61	
Db	23	ESDGET--KPLNNMGQICQICGDDVGALEGTGVDFAVACNECAFVCRPCRYEERKDDGTC	80	
OY	62	CPOCKTRRYRLKGSFPAIPDKDEDEGJADEGTYEFVNPCKEKISEFMLWMHLTRKGGEENG	121	
Db	81	CPOCKTRRRRRRSGSRVAGDEDEDV--DNIENEFYQA-----GANKAHQHRIG	128	
OY	122	EPQYDKEYSHNH--LPRLTSRQDTSGEFSAASPERLSVST--TAGSKRLPYSS--DV	173	
Db	129	E-ESSSSSRHSSQPIPLTLHGHVSGEI--RTPDQTSVRTSGPGLGPDNRMAISSPYIDP	185	
OY	174	NQSNRRRIYVD-----VGLGNVAMKEKVNDGCKMKMGKQENGTGVSVOAASERGVIDAST	227	
Db	186	ROPVPRVRLVSDSKDINSYGLANDVMDKEKEVEGKTKLOERNNMLQMTKYEHGKG--EIE--GT	243	
OY	228	DILADEALLNDENARQPLSRKKSIPSSRIINPYRMVIMLRLVILCLFHLRTIPVPAFAAL	287	
Db	244	GSNGEELQMAODTRLPMRSRVPIPSRSLPLRYVILLLIILICFLQYRTTHPVANAIFL	303	
OY	288	WLVSYICETWALSHILDQPKKWEPVNRETYLRLALRYDREGEPSQLAADVIFVSTVD	347	
Db	304	WLTSYICETWAFSLILDQPKKWPVINEFETLRLAIYDRDGEPSQLAADVIFVSTVD	363	
OY	348	LKEPPLVANTVLSLADVVDVKYSCVFPDGAAMLSFEETAEISFARKKVPCKKYS	407	
Db	364	LKEPPLVANTVLSLSDYVDVKACVSDGSSMLTFSEISTEAEPAKKWVPCKKFN	423	
OY	408	IEPAPAEVYFAAKIDYLTCKQVTSFVFKRRAMKREYEEFKIRINLVSFKALCKPREGWVM	467	
Db	424	IEPAPAEVYFAOKIDYLTCKQVTSFVFKRRAMKREYEEFKYRINALVAKQKIPREGWTM	483	
OY	468	ODGTPWPGNNTGDHPGMIQVFLGONGGLDAEGBNELPRLVYVSREKRPFGOHKKAGANNA	527	
Db	484	ODGTPWPGNNTRDHPGMIQVFLGSHGGIDTQGNELPRLIYVSREKRPFGOHKKAGANNA	543	
OY	528	LVRYSAVITLNPGLTINDCDHYINNSKSLREAMCGLMNPNGKQVCYQVQFORPFGIDKN	587	
Db	544	LVRYSAVITLNPAYLINDVCDHYFNNSSKSLIKKAMCMMPDPAIGKCCYQVQFORPFGIDILH	603	
OY	588	DRVARNRNTVFEDINLRGLDGIQGPVYVGTGCVFNFTALYGEPPYKTVHKKRPSLSKLCG	647	
Db	604	DRVARNRNTVFEDINMMKGLDGIQGPVYVGTGCCCFNRQALIXGDPVLTEEDLEPNITVKSQC	663	
OY	648	GSRRKNSKAKKESQKKSNG--RHTDSTVPEVNLDDIEEGVEGAGFDEKALLMSQMSLEKR	706	
Db	664	GSRRKKSQKSKYNEKREGRGIRNSDSNAPLFNMWEDIDEGFE--GYDDERSILMSQSVSEKR	721	
OY	707	FGQSAVPAFASHTLNGCVGVPASATENLKEAIAHYISCGYEDEKSDMGMEIGWYGVTEDI	766	
Db	722	FGQSPVFVLAATFEMOGGIPPTNPATLTLKEAIAHYISCGYEDEKTEGKIGWYGVSTEDI	781	
OY	767	LTGKEMHARGRSIYCMPLPAFGSAPAINLSDRLNOVLRLMAGVELTFSHHCYIWYG	826	
Db	782	LTGKEMHARGIYICNPAPPAFGSAPAINLSDRLNOVLRLMAGSITETLLSHHCYIWYG	841	
OY	827	NGRLKFLERAYVNTTIIYPIITSPLMCTLLAACLTFNOFIPIDISINIASIWEISLPLS	886	
Db	842	HGRRLRLERAIYNTIYVPIITSPLIACILPAPCLITDRFIIPISINYSWASITWITLFLIS	901	
OY	887	IFATGILBMRWSGVGIDEMWRNEQFWIGVSAHLFAVQGLKVLGADITNEFTVTSKAS	946	

Db	902	IAYGIIIELRMSQSI	IEDMMRNEDQFW	YIGCSAHLEAF	QGLLAVAGID	INFTNVT	TSKAT	961	
Qy	947	DEDGFAELIYFK	WTLLIIP	TTLLIYNLV	GVVAGV	ASTAIN	SGQSWG	PLFGKLEFFAFW	1006
Db	962	DEDGFAELIYFK	WTALLIIP	TTLLIYNLV	IGIVAGV	AAVNSG	QSWG	PLFGKLEFFAFW	1021
Qy	1007	IVHLIYPLFK	GLMG	RONR	PTI	VVWVS	YLLASIF	SLTLWRIDPF	1049
Db	1022	IATHLIYPLFK	GLMG	RONR	PTI	VIVWVS	YLLASIF	SLTLWRINPF	1064

ID	09LL14	PRELIMINARY:	PRT:	1059 AA.
AC	09LL14			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Cellulose synthase-6.			
DE	Cesa-6.			
OS	zea mays (Maize)			
OC	Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;			
OC	Panicoidae; Andropogoneae; Zea.			
OX	NCHI_TaxID=4577;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20398328; PubMed=10938350;			
RA	Holland N., Holland D., Helentjaris T., Dhugra K.S.,			
RA	Xocostle-Gazares B., Delmer D.P.;			
RT	"A comparative analysis of the plant cellulose synthase (Cesa) gene			
RT	family.";			
RL	Plant Physiol. 123:1313-1324(2000).			
DR	EMBL: AF200530; AAF89966.1; -			
DR	InterPro: IPR005150; Cellulose_synth.			
DR	InterPro: IPR001841; ZnF_ring.			
DR	Pfam: PF03552; Cellulose_synth. 1.			
DR	SMART: SM00184; RING: 1.			
SO	SEQUENCE 1059 AA; 120044 MW; 289DA26B252322249 CRC64;			
Query Match	68.7%; Score 3900; DB 10; Length 1059;			
Best Local Similarity	67.5%; Pred. NO.3.9e-294;			
Matches 728; Conservative 140; Mismatches 154; Indels 56; Gaps 17;				
OY	18 QTCQICSDNKGTVGDGRFVACDICSPFVRCPCYERKRDGNOSQPOCKTRKRLKSGPA 77			
DB	7 QVCOICGDGVGRNPDGEFVACNCAFCICRDCEYERRETQNCPOCKTRKRFKFGCAR 66			
OY	78 IRGKDEGLDAGEVFNYPQK--EKISRMRGMHLTRKGEEM-GEPIDDEVSHNH 133			
DB	67 VFGDEEGEGV-DLEENEFNMSDKHDSOTLASMALAHMSYRGADLDCVDPDFHPIP--N 123			
OY	134 LPRLTSRQ--DTSGEFSASPERLSVSTIAGKR--LPSSDVNSQPNKRIYDP-- 184			
DB	124 VPLLTNGQWVDIPPDQHALVP-----SFVGGGGRIRHLPY-ADPNLPVQPRSMDFSKD 177			
OY	185 ---VGLGVAMKKEVDDGKMKQEKNTGVSVOASERGVQVLDASTDLADPALLNDGAR 241			
DB	178 LAAYGYSGVAKKERESKQKQER---MHOTNRDGGGDDGD--ADPLM-DBAR 226			
OY	242 QPLSRKVSIPSSRIINPYRMVIMLVLVILFLYHRTIPVYNAPALMVSVCIEIWFALS 301			
DB	227 QPLSRKIDLPSSQINPYRMIIILVYLCEFFHRYVMHPVPDAPAMLIISVCEIWFAMS 286			
OY	302 WILDFPKWPFVNRBTYLDRLALRYDREGPESQLAADVIVSYVDPLKEPPLVANTVLS 361			
DB	287 WILDFPKWPFIERETYLDRSLSTLEFDGEGHSQLAPODFEYSTVDPLKEPPLVANTVLS 346			
OY	362 ILADVDPDKSCVVPFDGAMLSEESLAETSEFARKVVPCKKYSTEPRAPEMYFAKI 421			
DB	347 ILADVDPDKSCVSDGAMLTFEALSETSEFAKKVVPCKRISLEPRAPEMYFOOKI 406			



OY	422	DYLDKQVTSVVKORRAMKREYEERIKRINALVSALKCPBEGVMODGTPWGNNGDH	481
		: : : : :           : : : : :           : : : : :	
Db	407	DYLDKVAPNVRERRRAKKREYEERFKRINLVAKQVPEEGTMDGTPEPNNVDH	466
OY	482	PGMIQVFLGGONGGLDAEGNELPRLYUVYSREKRPGOHNRKAGANMALRVSAVLTNNGPI	541
		: : : : :           : : : : :           : : : : :	
Db	467	PGMIQVFLGGSGHDVEBNELPRLYUVYSREKRPGNHHRKAGANMALRVSAVLTAAYL	526
OY	542	LNLDCDHYNNSKALREAMCFILMPDNLGKQVCYOVPQRFPGIDIKNDYVARNRNVFFEDIN	601
		: : : : :           : : : : :           : : : : :	
Db	527	LNLDCDHYNNSKAIRKEMCFMMDPLRGKKCYOVPQRFPGIDRHDXARNRVFFEDIN	586
OY	602	LRGDLGDIGPYVYGTCVFNTALXGEPPRLTKVHKRPS-----LSLKCGSRK	651
		: : : : :           : : : : :           : : : : :	
Db	587	MKGIDGIDGPYVYGTCVFRRQAOLXYGDAF--RTKKRPSTRCCMCWPWCICCCCFGRNK	643
OY	652	KNSRAKKEESDK----KKSGRHTDSVPNFNDDIEEGVGSGFDERAKLLMSQSLEKRF	707
		: : : : :           : : : : :           : : : : :	
Db	644	TKKTKTISKPFEEKIKLFKKKKENQAPRYALGEIDEAARGA--EREKASTYNQOKLEKF	701
OY	708	GOSAVFYASTLMENGVPSPATPENLLKEALIHVISCEGEDSDMGMEIGMTYGSYTEIDL	767
		: : : : :           : : : : :           : : : : :	
Db	702	GQSSVFYASTLLENGTLKASASPALSLEKALHVHSISCEYEDKTGMCKDGMTYGSTIEDIL	761
OY	768	TGFPMHARGWSIYCMKPLPAFKSAPINISDRLOVRLMALGSEYLFSRRCPIWGYN	827
		: : : : :           : : : : :           : : : : :	
Db	762	TGFMHCHGMRSIYCIPKRAAFKKSAPLNLSDFHQVRLMALGSTIELFSNNHCPMYGYG	821
OY	828	GRLEPLEFAVVNTTIYPTISPIPLYCTLLAVCLFTNQFIITPOLSNASTIFSLSTFI	887
		: : : : :           : : : : :           : : : : :	
Db	822	GGLEPFESESTINSIVPWITSPIPLACTPLAICLLTGKFTTPELNNVASLMPMSLFICI	881
OY	888	FATGILERMRAGSGVIDEWMRNDEOFVIGVSAHLFAVPQGILUKVLADJTFNFTYSKASD	947
		: : : : :           : : : : :           : : : : :	
Db	882	FATSLERMRMSGVGDGMWRNEQVRIQGVSHLPFAVQGLLKVIAGVDTSFITYTSKGD	941
OY	948	EDGDFAEILYFERKWTLLIPTTLLIVNLGVAVAGVSAINSGYOSMBPLFGKLFPAFVI	1007
		: : : : :           : : : : :           : : : : :	
Db	942	-DEEFSELTYETKWTLLIPTTLLILNFIGVAGISAINNGYESMBPLFGKLFPFAFVI	1000
OY	1008	VHLXPFLKGLMGRNRPPTIYVWSVLLASFSLNVAVIDPFTSVYTPDILEGCINC	1065
		: : : : :           : : : : :           : : : : :	
Db	1001	VHLXPFLKGLVGRNRPPTIYVWSILLASFSLNVAVIDPFLAKDDPDLLECGLDC	1058
<hr/>			
RESULT 11			
O9LLI3			
ID	O9LLI3	PRELIMINARY;	PRT; 1086 AA.
AC	O9LLI3:		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Cellulose synthase-7.		
CN	CESA-7.		
OS	Zea mays (Maize).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; Liliopsids; Poales; Poaceae; PACC clade;		
CC	Panicoidae; Andropogoneae; Zea.		
OX	NCBI_TaxID=4577;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20398328; PubMed=10938350;		
RA	Holland N., Holland D., Helentjaris T., Dhugga K.S.,		
RA	Xocoostle-Cazares B., Delmer D.P.,		
RT	"A comparative analysis of the plant cellulose synthase (Cesa) gene		
RT	family,"		
RL	Plant Physiol. 123:1313-1324(2000).		
DR	EMBL; AF200531; AAF89967.1; -		
DR	InterPro: IPR005150; Cellulose_synth.		
DR	InterPro: IPR001841; znf_ring.		
DR	Pfam; PF03552; Cellulose_synth; 1.		
SMART; SMO0184; RING; 1.			
SO	SEQUENCE 1086 AA; 122608 MW; DD03C73ABDJ3EZEB CRC64;		

Query Match	68.3%;	Score 3875.5;	DB 10;	Length 1086;
Best Local Similarity	66.7%;	Pred. No. 3.2e-292;		
Matches 728;	Conservative 145;	Mismatches 161;	Indels 57;	Gaps 18;

QY	4	4GEFAGKRMKNIVPOTQOICSDNWKYTDGDRFVACDICSIPVRCPCYEKEKQDNOSCP	63
QY	23	DGDDEPKPRRQNGVGVICGDDVDGLAGDGFVACNCAFPVVCXDCEYEKEBETQWCP	82
QY	64	QCKTRKYRKLKSSPAIPGDKDEGLADECTVEFNYP--QKEKISEBMLQWHLTRGK--GEEM	120
Db	83	QCKTRKYRKLKCOQARTGDEBEDGV--DDLNDENMNMGHDSQSASBMLGHMSYGGSDPN	141
QY	121	GEPOYDKEVSHNHLEPRLTSRQ---DTSEFSASPERLVSSTJAGKR--LPSYSDVN	174
Db	142	GAPQAFQ--LNPVNPVLITNGOMVDIPEQOHALVPSFMG---GGGRIRHLPX--ADPS	193
QY	175	QSPNRIYDP-----VGLNVAMKERDWMKQEKNTGVPSTQAASERGGVIDASTD	228
Db	194	LPVQDRSPMDKDLAAYGVSVAMKERMENMKORER-----MHQTDNGDGGDDDD--	245
QY	229	ILADEALLNDEAROPLSRKRVSIPSSRINPYPVAVMLRLVILCFLHRYITNPVPAPALM	288
Db	246	--ADPLPM--DEARQOLSKRIPLPSSQINPYRMIIIRLVLVGFFHYVMHVPNDAPALM	302
QY	289	LVSVCIEIWFALSWLDOFPKWFNPVNETYLDRLALRYDRGSEPSQLAVIDFVSTVDP	348
Db	303	LIVSICEIWFAMSWLDOFPKWFPIERETYLDRLSLRDKBEGOPQLAPIDFVSTVDP	362
QY	349	KEPPLVTANTYLSILAADVVDKVCYCFDDCAAMLSPESLAETSEPAKRWVPCKKXSI	408
Db	363	KEPPLVTNTYLSILSVDPYDKVSCYSDDCAAMLTEALSETSEPAKRWPFCKRINI	422
QY	409	EPRAPEWYFAAKIDYLDKRVOTSEFYKDRBRAMKREYEERKIRINALVSKALCPBEGWYQ	468
Db	423	EPRAPEWYFOQKIDYLDKQVANAIFYREBRAMKREYEERKVRINALVANAQVPEBGWYQ	482
QY	469	DGTPPFGNNTGDHPMIQVFLGQNGGLDAEGNELPRLVYVSSEKRPQOHKKAAAMAL	528
Db	483	DGTPPFGNNVDRHPMIQVFLGQSGGLDCEGNELEPRLVYVSSEKRPQYHNHKKAAAMAL	542
QY	529	VRSVAVLNGPRLILDCDHYIINNSKALREALCFMLDNLGQVOCYVOFPORFQDIDKND	588
Db	543	VRSVAVLNLAPRLILDCDHYIINNSKALKEAMCFMDELPLGKVCYVOFPORFQDIDKND	602
QY	589	RYANRNTYFEDNLGDLGQGVYVGTGVFNRTALGYEBPIKVKHKKPSLSKLG-	647
Db	603	RYANRNVYFEDINMGGLDGIQGPVYVGTGVFRQALGYDAP--KTKKRP--SRTQNC	657
QY	648	-----GSRKKNSKAKKESKRR--SGRHTDSVYVPVNLDDIEBGVAGCFDDEK	694
Db	658	WPKWCFCCCCGNGNRQKTKRPTKTEKKRLLEFKKEENOSPAVALGEIDEAEGA--ENEK	715
QY	695	ALLMSOMSLERFGGSAFVASTLMENGVPSPAPENLKALHIVISGQYDKSDMGME	754
Db	716	AGIVAOQLEKKFGOSSVYFVSTLENGSTLKSASPAStLKPAIHIVISGQYDKSDMGKE	775
QY	755	IGWIGSVTEDLITGFKMHARGWRSIYCMPLKPAERGSAPINLSDRLNOVLRMALGSVEI	814
Db	776	IGWIGSVTEDLITGFKMHCHQMRSIYCIKPKVAFKGSAPINLSDRLNOVLRMALGSJEI	835
QY	815	LFSRRCPIWYGYNGKLFLEPFAVYNTIITYPTISPLMKYCTLAVCFITNQFIIPQSN	874
Db	836	FFSNRCPIWYGGGLKFLERSTYNSIYVPTISPLLAYCTLPALICLTGTGFIPELNN	895
QY	875	IASIFLBSLIFSTAGLIERMWSGVGIDEMWRNRQFNVIGVSAHLFAVEQGLIKVLAVG	934
Db	896	VASLMEFNLFCIFATSLIERMWSGVGIDEMWRNRQFNVIGVSHLFAVFOGLIKVLAVG	955
QY	935	IDTNEFTVSKASDEDEGFAELYLFKRWTLILPPTTLILVNLGVAVGVSAVINSQYQSWG	994
Db	956	VDTSFTVYVSKGCD--DEEFSLEYTFKWTLLILPPTTLILNLTGVAGVSAVINSQYQSWG	1014

```

QY 995 PLFGKLEFAFWIYVHLVPLKGLMGGRNPTIIVVSVLLASIFSLLWVRIDPFTSRVY 1054
DB 1015 PLFGKLEFAFWIYVHLVPLKGLMGGRNPTIIVVSVLLASIFSLLWVRIDPFLAKD 1074
QY 1055 GPDILECGINC 1065
DB 1075 GPLLECGDLC 1085

RESULT 12
09LL12 PRELIMINARY; PRT: 1094 AA.
AC 09LL12;
DB 01-OCT-2000 (Tremblrel, 15, Created)
DB 01-OCT-2000 (Tremblrel, 15, Last sequence update)
DB 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Cellulose synthase-8.
GN CESA-8.
OS Zea mays (maize).
OC Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Eukaryota; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20398328; PubMed=10938350;
RA Holland N., Holland D., Helentjaris T., Dhuga K.S.,
RA Xocoostle-Cazares B., Delmer D.P.;
RA "A comparative analysis of the plant cellulose synthase (cesa) gene
RT family.";
RL Plant Physiol. 123:1313-1324(2000).
DR EMBL; AF200532; AAF89968.1; -
DR InterPro; IPR001841; Znf_ring.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF03552; Cellulose_synth; 1.
DR SMART; SM00184; RING; 1.
SQ SEQUENCE 1094 AA; 122575 MW; DFBBC18F49D23F5E CRC64;

Query Match 67.7%; Score 3843.5; DB 10; Length 1094;
Best Local Similarity 66.3%; Pred. No. 1e-289;
Matches 727; Conservative 134; Mismatches 178; Indels 57; Gaps 17;

QY 2 ESEGTACKPMKNIVPOTCISDNVGTVDGDRVYACISFPVPCRYEYERKDNOS 61
DB 23 DESGSAAGGAARRAARPCOICGEVGFDEPFVACNECAFPCVACRYERREGSOA 82
QY 62 CPOCKTRVRLKGSALPDGKDEGLAD-----EGVEENYPOKEKISERMLGMLT 113
DB 83 CPOCKTRVRLKGSALPDGKDEGLAD-----EGVEENYPOKEKISERMLGMLT 113
QY 114 RKGGEEMGEPOYDKESVSHNLRLTSRQ---DTSGEFSASPERLSVSSTIAGKR--L 167
DB 141 YRGGD-AHGRSPV---PNVPLTLNGCWVDDIPRQHLALVSYNGGG--GGKRIHPL 194
QY 168 PVSVDVNSPNRIYDP-----VLGNVAKERVDGKMKOEKNTGPVSTOASERGV 221
DB 195 PF-ADPNLNFVQPSMDPSKDLAAYGVAVAKERMGKOEKOEK-----LQHVSEGGG 247
QY 222 DIDASTDLADALINDEAROPLSRKVSISSIRINPVRVAVIIVLTLFLHYRTINV 281
DB 248 DMGDGD--ADLPLM-DEAROPLSRKVSISSIRINPVRVAVIIVLTLFLHYRTINV 303
QY 282 PNAFLMLVSVICELIFALSWILDOFPKMFVYNRETYDLRLALRYDEGEPSQLAAVDIF 341
DB 304 KDAFLMLVSVICELIFALSWILDOFPKMFVYNRETYDLRLALRYDEGEPSQLAAVDIF 363
QY 342 VSTVPLKEPPLVNTAVSILAVDPVDKVCYVFDGAAMLSEFSAETSEFAKWP 401
DB 364 VSTVPLKEPPLVNTAVSILAVDPVDKVCYVFDGAAMLSEFSAETSEFAKWP 423
QY 402 FCKKYSIEBRAEWFAKIDYLDKQVQTSFYKDRAMKREVEEFIRINALYSALKEP 461
DB 424 FSKKNIEBRAEWFAKIDYLDKQVQTSFYKDRAMKREVEEFIRINALYSALKEP 483

```

```

QY 462 EEWVMDQGTQPMWNTGDPGMIQVFLQNGSLDAEGNELPRLVYSEKRRSEFOHKK 521
DB 484 EEWVMDQGTQPMWNTGDPGMIQVFLQNGSLDAEGNELPRLVYSEKRRSEFOHKK 543
QY 522 AGANALVAVSALVINGPILINDCHYINNSKALREACFLMDPMLGVOCVYOPOR 581
DB 544 AGANALVAVSALVINGPILINDCHYINNSKALREACFLMDPMLGVOCVYOPOR 603
QY 582 DGDIDKNDRYANRVTFEDINLNGLOSIGPPVYVGTGCVNRNALGYEPPYKHKPPS 640
DB 604 DGDIDKNDRYANRVTFEDINLNGLOSIGPPVYVGTGCVNRNALGYEPPYKHKPPS 660
QY 641 -----LSTRKLC--GGSRRKNSKAKKESPKKSGRTHDSTVPFNLDIEGVEGAG 689
DB 661 RHCNCPKMKLSCCCKRNNKKTTPKTEKKRKLFEKKAENSPAYALGEIDEGAPGA- 719
QY 690 FDEKALMSQMSLEKRGQSAVFAVASTLMENGVPSPATPENLKEAIVHISGCEYDXS 749
DB 720 -DIEKAGIYNOCKLEKKEKFGQSSVFAVASTLLENGTLKSASPASLKEAIVHISGCEYDXS 778
QY 750 DMGMEIGWYGSVTEIDILTFKMHARGMSIYCMPLPAFKGAPINLSDRLOVLRMAL 809
DB 779 DMGMEIGWYGSVTEIDILTFKMHARGMSIYCMPLPAFKGAPINLSDRLOVLRMAL 838
QY 810 GSVEILFSRHCPIWYNGRLKFLERFAVNTTYPITSIPILMYCTLLAVCLFTNOFTI 869
DB 839 GSVEILFSRHCPIWYNGRLKFLERFAVNTTYPITSIPILMYCTLLAVCLFTNOFTI 898
QY 870 POISNATSWLSLFLSATGILKEMRSGVGDENRNEQFWIGVSAHLEPAVFOGIL 929
DB 899 PELTNVAVSIFMAFLCISVGTGLEMRSGVAVIDDWRNEQFWIGVSAHLEPAVFOGIL 958
QY 930 KYLAGIDTNTFTYTSKASDGDGFALYLLKMTLLIPPTLLIVNLGVAGSVYAINSG 989
DB 959 KVPAGIDTNTFTYTSKASDGDGFALYLLKMTLLIPPTLLIVNLGVAGSVYAINSG 1017
QY 990 YQSWGPLFGKLEFAFWIYVHLVPLKGLMGGRNPTIIVVSVLLASIFSLLWVRIDP 1049
DB 1018 YQSWGPLFGKLEFAFWIYVHLVPLKGLMGGRNPTIIVVSVLLASIFSLLWVRIDP 1077
QY 1050 TSBVTPGPDILECGINC 1065
DB 1078 LAKSNGLPDLCEGLDC 1093

RESULT 13
09SWM6 PRELIMINARY; PRT: 1026 AA.
AC 09SWM6;
DB 01-MAY-2000 (Tremblrel, 13, Created)
DB 01-MAY-2000 (Tremblrel, 13, Last sequence update)
DB 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Cellulose synthase catalytic subunit.
GN IRX3 OR TI0B6_80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Magnoliophyta; Magnoliopsida; core eudicots; Rosidae;
OC eudicots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. LANDSBERG ERRECTA;
RA MEDLINE=99264300; PubMed=10330464;
RA Taylor N.G., Scheible W.R., Cutler S., Somerville C.R., Turner S.R.;
RA "The irregular xylem3 locus of Arabidopsis encodes a cellulose
RT synthase required for secondary cell wall synthesis.";
RL Plant Cell 11:769-780(1999).
[2]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RA Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.

```

```

RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF091213; A040885.1;
DR EMBL: AL391142; CAC01737.1;
DR InterPro: IPR001510; Cellulose_synth.
DR InterPro: IPR001841; znf_rling.
DR Pfam: PF03552; Cellulose_synth; 1.
DR SMART: SM00184; RING; 1.
SQ SEQUENCE 1026 AA; 115797 MW; 503BFBCT8BE6E511 CRC64;

Query Match      67.1%; Score 3811; DB 10; Length 1026;
Best Local Similarity 66.9%; Pred. No. 3.1e-287;
Matches 721; Conservative 115; Mismatches 143; Indels 98; Gaps 17;

QY 10 KPMKIVIVPOTQICSDNKGTVYDGRFVACDICSPPVCPCEYERKKGDNOSCPOCKTRY 69
DB 27 KPLKMLDQFCETICDQIGLVEGDLFVACNCGCPACRCPCEYERREGTONCPOCKTRY 86
QY 70 KRLKGSAPAI PGDKDEDGLAD---EGTVEFNYPQKKEISERMIGMHLTGKGEKME-POY 125
DB 87 KRLKGSAPVPEGEDEDEDIDIEYFNIEHEDDKRHSAEAMLYGKMSYGRGPEDENGRF 146
QY 126 DKEYSHNLPRILTSQDTSGEFSASPERLSVSTIAGG-----KRL-PPSSDVNQ 175
DB 147 PPVYAGH-----SGEF-----PVGGYGNGEHLKRVHPYSSENG 184
QY 176 SPNRITVPVGLGNVAMKERVDMKMKOEKNTGPVSTOASERGVVIDASTDIIADLEAL 235
DB 185 SEG-----GMRERMDMKL-QHGNLGPPEP-----DDDEPKGLI----- 216
QY 236 LNDEAROPLSRKVSIPSSRINPYRMVIMRLVILCLFLHYRTNPVAPAFALMYSVCE 295
DB 217 --DEAROPLSRKVPASSKINRYRNVIAVLAVFLKRLNVLNVDHALGMLTSVCE 274
QY 296 IMFALSWILDQFPKMFVNRETYDLRLALRYDREGEPSQLAAVDFVSTVDPKPPLYT 355
DB 275 IMFALSWILDQFPKMFIERETYLDRLSLREREGEPNMLAPVDVFSVDPKPPLYT 334
QY 356 ANTVLSIADVIVDVKVSCYFYDDGAMLSFEBSLAETSEFARKVWPFCKKYSIEPRADW 415
DB 335 SNTVLSIADMPYVPEKISCYVSDGASMLTESLSETAEFARKVWPFCKKFSIEPRADW 394
QY 416 YFAAKIDYLDKVOFSYFKDRAMKREYEEFKIRINALYSKALKCPREGVWVQDSTPMPG 475
DB 395 YFTLVVDYLDQVHFTFVKERRAMREYEEFKVRINAOVAKSKVPLEGWLMDGTMPMG 454
QY 476 NNTGDHPGMIOVFLGONGGLDAEGNELPRLVYVSREKRPFGQHHKAGAMNALVVSAYL 535
DB 455 NNTKDPGMIOVFLGSHSGFDEVEGHELPRILVYVSREKRPFGQHHKAGAMNALVVSAYL 514
QY 536 TNGPPIILDCDHTYNNKALREAMCFIADPNLGAOVGYOEPORFQDIDKNDRYANRNT 595
DB 515 TNAPMLMLDCDHTYNNKALREAMCFIADPNLGAOVGYOEPORFQDIDTDRYANRNT 574
QY 596 VEFEDNLGDLGDIQGVYVGTGVENRTALGYEPRIVKHKPSLLSKLGS-----GSR 650
DB 575 VEFEDNLGDLGDIQGVYVGTGVENRTALGYEPRIVKHKPSLLSKLGS-----GSR 630
QY 651 KKNKSAKKSDESKKSGRHTDSTVPVFNLDIEEGVEG-AGFDEKALMSQMSLEKRFQ 709
DB 631 RKNKK-----FSKNDNMGDVAALAGAGDEKHLMSKNFKETGQ 670
QY 710 SAVFASVTLMEGVPSPATPENLKEAIIHVSCTYEDKSDMGMTIGMYSVEDIILTG 769
DB 671 SSIFATSTLMEGVPSPSSPAVYLKEAIIHVSCTYEDKTEWGTGLMGTIGITDIIITG 730
QY 770 FKMHAQRSTIYCMPLKPAFKGSAPINLSDRLNOVLRLMALGSVELLFRRHOCIMWY-NG 828
DB 731 FKMHAQRSTIYCMPLKPAFKGSAPINLSDRLNOVLRLMALGSVELLFRRHOCIMWY-NG 790
QY 829 RLKFLERPAVYNTIYPTISIPILMYCTILAVCLFTNQFIIPQISNINIASIMFLSLISIF 888

```

```

DB 791 KKLWLERPAVYNTIYPTISIPILMYCTILAVCLFTNQFIIPQISNINIASIMFLSLISIF 850
QY 889 ATGILEMRNSGVGIDEMWRNEQFWYIGVSALHFAVFGIILKVLGIDIONFVTSKASDE 948
DB 851 VTGILELRWSGVSIEMWRNEQFWYIGVSALHFAVFGIILKVLGIDIONFVTSKADDD 910
QY 949 DGDFAELVLFKWTTLIPPTTLIYNLVGVGVSAIYNSQSGMPLFGKLFPAFWYI 1008
DB 911 D-DGSELVAFKKTTLIPPTTLIYNLVGVGVSAIYNSQSGMPLFGKLFPAFWYI 969
QY 1009 HLYPLKGLMGQRNPTPIVWVSVLASISFSLVWRIDPFTSRVGPDIIECGINC 1065
DB 970 HLYPLKGLMGQRNPTPIVWVSVLASISFSLVWRIDPFTSRVGPDIIECGINC 1026

RESULT 14
Q9XHP6 PRELIMINARY; PRT; 1026 AA.
ID Q9XHP6:
AC Q9XHP6:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cellulose synthase catalytic subunit.
GN IRX3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV LANDSBERG, ERECTA.
RX MEDLINE=99264300; PubMed=10330464;
RA Taylor N.G., Schelble W.R., Cutler S., Somerville C.R., Turner S.R.;
RT "The irregular xylem3 locus of Arabidopsis encodes a cellulose
  synthase required for secondary cell wall synthesis.";
RL Plant Cell 11:769-780(1999).
DR EMBL: AF089917; AAD32031.1;
DR InterPro: IPR001510; Cellulose_synth.
DR InterPro: IPR001841; znf_rling.
DR Pfam: PF03552; Cellulose_synth; 1.
DR SMART: SM00184; RING; 1.
SQ SEQUENCE 1026 AA; 115858 MW; 453BRDID283C4D70 CRC64;

Query Match      67.0%; Score 3805; DB 10; Length 1026;
Best Local Similarity 66.9%; Pred. No. 9e-287;
Matches 720; Conservative 115; Mismatches 144; Indels 98; Gaps 17;

QY 10 KPMKIVIVPOTQICSDNKGTVYDGRFVACDICSPPVCPCEYERKKGDNOSCPOCKTRY 69
DB 27 KPLKMLDQFCETICDQIGLVEGDLFVACNCGCPACRCPCEYERREGTONCPOCKTRY 86
QY 70 KRLKGSAPAI PGDKDEDGLAD---EGTVEFNYPQKKEISERMIGMHLTGKGEKME-POY 125
DB 87 KRLKGSAPVPEGEDEDEDIDIEYFNIEHEDDKRHSAEAMLYGKMSYGRGPEDENGRF 146
QY 126 DKEYSHNLPRILTSQDTSGEFSASPERLSVSTIAGG-----KRL-PPSSDVNQ 175
DB 147 PPVYAGH-----SGEF-----PVGGYGNGEHLKRVHPYSSENG 184
QY 176 SPNRITVPVGLGNVAMKERVDMKMKOEKNTGPVSTOASERGVVIDASTDIIADLEAL 235
DB 185 SEG-----GMRERMDMKL-QHGNLGPPEP-----DDDEPKGLI----- 216
QY 236 LNDEAROPLSRKVSIPSSRINPYRMVIMRLVILCLFLHYRTNPVAPAFALMYSVCE 295
DB 217 --DEAROPLSRKVPASSKINRYRNVIAVLAVFLKRLNVLNVDHALGMLTSVCE 274
QY 296 IMFALSWILDQFPKMFVNRETYDLRLALRYDREGEPSQLAAVDFVSTVDPKPPLYT 355
DB 275 IMFALSWILDQFPKMFIERETYLDRLSLREREGEPNMLAPVDVFSVDPKPPLYT 334

```

QY 356 ANTVLSILAVDPVDCVYEDDGAAMLSESLAETSEFARKVWPFCKKSTIEPRAPEM 415  
 DB 335 SNTLSILAMDYEVEKISCVSDDGASMLTFESLSAEPRARKVWPFCKKSTIEPRAPEM 394  
 QY 416 YFAKIDLYLKDQVQTSFVDRBRAMKREVEEFKIRIINALSKAKCEBEGVMDGTPMNG 475  
 DB 395 YFTLKVADYLDQKHPFVKEERRAMKREVEEFKIRIINAQAKASKVPLEGIMDGTWPWG 454  
 QY 476 NNTGDHPRGMIOVFLGONGGLDAGNELPRLVYVSREKRGFOHKKAGANALVRSAYL 535  
 DB 455 NNTKDHPRGMIOVFLGSGGDFVGHLPRLVYVSREKRGFOHKKAGANALVRSAYL 514  
 QY 536 TNGPFLNLDCOHYINNSKALRPAKCFGLADPNLCKOVYVQFQFQDIDKNDVYANRNT 595  
 DB 515 TNAFPMNLNDCOHYVNSKAVREAKCFGLADPNLCKOVYVQFQFQDIDTNDVYANRNT 574  
 QY 596 VFEDINLRGLDGIQGVYVGTGCVFNRTALYGEPRIPKVKHKKPSLSLKCG-----GSR 650  
 DB 575 VFEDIMKGLDGIQGVYVGTGCVFNRTALYGEPRIPKVKHKKPSLSLKCG-----GSR 630  
 QY 651 KNSKAKKESDKKKSGRHTDSTVPFNLDIEGVEG-AGFDEKALMSOMSLERFQ 709  
 DB 631 RKNKK-----FSKNDMNGDVALGAEKDEKHLMEFEMFETFG 670  
 QY 710 SAVFASTLMENGVPSPATPENLKEAIVHISCGEDSKDMCEIGMIGYSTEDILTG 769  
 DB 671 SSIFVSTLMENGVPSPSPAVLKEAIVHISCGEDSKDMCEIGMIGYSTEDILTG 730  
 QY 770 FKHHAGWRSIYCMPLPAFGKSPINLSRLNQVRMAIGSEILFHSRCHPIWGY-NG 828  
 DB 731 FKHHAGWRSIYCMPLPAFGKSPINLSRLNQVRMAIGSEILFHSRCHPIWGY-NG 790  
 QY 829 RLKFLERFAYVNTIYPTISPLMYCTLLAVCLFTNOPIQOISNIASIMPLSLFISF 888  
 DB 791 KTKMLERFAYVNTIYPTISPLMYCTLLAVCLFTNOPIQOISNIASIMPLSLFISF 850  
 QY 889 ATGILMRNSGVIDEMWNRQWVIGVSAHLFAVFOGILKVLADIDNFTYTSASDE 948  
 DB 851 VTGILMRNSGVIDEMWNRQWVIGVSAHLFAVFOGILKVLADIDNFTYTSASDE 910  
 QY 949 DCGFALYLFKMTLLIPPTLLIVLVYAGVSAVNSGVOSWPGFEGKLFEPANVIV 1008  
 DB 911 D-DEGLYAFKMTLLIPPTLLIVLVYAGVSAVNSGVOSWPGFEGKLFEPANVIV 969  
 QY 1009 HLYPFLKGLMGRNRPPTIVVWVSLASIFSLWVRIDPFTSRVTPDILEGGINC 1065  
 DB 970 HLYPFLKGLMGRNRPPTIVVWVSLASIFSLWVRIDPFTSRVTPDILEGGINC 1026

RESULT 15  
 09SKJ5 PRELIMINARY; PRT; 1065 AA.  
 AC 09SKJ5;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Putative cellulose synthase catalytic subunit.  
 GN ATG25540.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBITaxID=3702;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shoa T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Meson T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Coppenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC006300; ADD20713.1;  
 DR InterPro: IPR005150; Cellulose\_synth.  
 DR InterPro: IPR001309; ICE\_p20.  
 DR InterPro: IPR001841; ZnF\_Cing.  
 DR Pfam: PF03552; Cellulose\_synth; 1.  
 DR SMART: SM00184; RING; 1.  
 DR PROSITE: PS01121; CASPASE\_HIS; UNKNOWN.1.  
 SQ SEQUENCE 1065 AA; 120616 MW; F2AE36C9492A6076 CRC64;

Query Match 67.0%; Score 3801; DB 10; Length 1065;  
 Best Local Similarity 66.2%; Pred. No. 2e-286;  
 Matches 714; Conservative 132; Mismatches 156; Indels 76; Gaps 12;

QY 10 KPMKNVPTQICSDNNGKTYDGRFVACDICSFPVCPYERKBDGNSQPOCKTAY 69  
 DB 23 KPLKDLNGOICQCGDVGTLTGTVNVAACNCGFPLCSCYTEERKDSQCPCKANF 82  
 QY 70 KRLKSPALPGKDEKDEGLADEGVTFENYPOKEKISERMLGMHLTRNGGEMGEPOYDEY 129  
 DB 83 RRLNSPRVREVEKEDVDVNDIEN-EDFYQ----- 111  
 QY 130 SHNHLPRLTSRODTSGEFSAAAS--PERLSVSTIAGKRLPYSSDVNOSPNT----- 179  
 DB 112 -GNKKARLPRAE---EFSSSRHRESLFLVSLTHGH--PVSGET-PPPDHATLSPCI 163  
 QY 180 -----RIVP-----VGLGNVAKRVDGMKMKOENKGPYSTOASBR 218  
 DB 164 DPQLPEITVOLLPLVRLIDPSKDLNLSYGLVNDMKRIQGMKLDKQKNHMTGKHBEK 223  
 QY 219 GGVYDIDASTDIIADENALNDEARQPLSRKVSIPSRINRYRMVIMLRVILDLFLHYRT 278  
 DB 224 GGE--PEGISNGDELQWVDARLPMSRVVHPSPAMPTRYRIVYRLILDLVFLHYRT 281  
 QY 279 NPYNAPFALMLVSYICEIFALSMIILDOPPKFPPVAKRETYLDRLALRYRESEPSQLAAY 338  
 DB 282 HPVKDAVYALMLTSVICEIWFASFMLDOPPKYPIRRETFLLRLALRYRDEEPSQLAAY 341  
 QY 339 DIFVSTVDLKEPPLTANTVLSILAVDPVDCVYEDDGAAMLSESLAETSEFARKV 398  
 DB 342 DVFVSTVDLKEPPLTANTVLSILAVDPVDCVYEDDGAAMLSESLAETSEFARKV 401  
 QY 399 WPFCKKSTIEPRAPEMVYFAKIDLYLKDQVQTSFVDRBRAMKREVEEFKIRIINALSKAL 458  
 DB 402 WPFCKKSTIEPRAPEMVYFAKIDLYLKDQVQTSFVDRBRAMKREVEEFKIRIINALSKAL 461  
 QY 459 KCPBEGWVMDQTPWFGNNTGHPGMIOVFLGONGGLAEGNDELPRLVYVSSEKRGFOH 518  
 DB 462 KIPBEGWVMDQTPWFGNNTGHPGMIOVFLGONGGLAEGNDELPRLVYVSSEKRGFOH 521  
 QY 519 HKKAGANALVRSAYLNGPFLNLDCOHYINNSKALRPAKCFGLADPNLCKOVYVQFQF 578  
 DB 522 HKKAGANALVRSAYLNGPFLNLDCOHYINNSKALRPAKCFGLADPNLCKOVYVQFQF 581  
 QY 579 QRFQDIDKNDVYANRNTVFEDINLRGLDGIQGVYVGTGCVFNRTALYGEPRIPKVKHKK 638  
 DB 582 QRFQDIDKNDVYANRNTVFEDINLRGLDGIQGVYVGTGCVFNRTALYGEPRIPKVKHKK 641  
 QY 639 PSLSLKLGSGSRKNSKAF-KESDKKSGRHTDSTVPFNLDIEGVEGAEFDEKALL 697  
 DB 642 PNIVYSCGSGSRKNSKAF-KESDKKSGRHTDSTVPFNLDIEGVEGAEFDEKALL 699  
 QY 698 MSQMSLEKFGSGSAFVASTLMENGVPSPATPENLKEAIVHISCGEDSKDMCEIGM 757  
 DB 700 MSQMSLEKFGSGSAFVASTLMENGVPSPATPENLKEAIVHISCGEDSKDMCEIGM 759



THIS PAGE BLANK (USPTO)

T  
..... (USPTO)